

Essays on the Real Economy and the Financial Sector

Dissertation
submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktorin der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Sabrina Studer
from St.Gallen, SG

approved in February 2016 at the request of

Prof. Dr. Dr. Josef Falkinger
Prof. Dr. Fabrizio Zilibotti

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 17.02.2016

Chairman of the Doctoral Board: Prof. Dr. Steven Ongena

Acknowledgements

Writing a PhD thesis is a process which cannot be accomplished without substantial scientific and personal support. While writing my dissertation at the chair Public Finance and Macroeconomics at the Department of Economics of the University of Zurich I benefited a lot from all the people around me:

I will always be most indebted to my supervisor Josef Falkinger. I thank him for the profound guidance, the dedicated mentoring and all the support and motivation I experienced from him over the last years. His advices taught me to conduct research in a rigorous way and the innumerate discussions with him triggered my economic curiosity. Beyond that, it was very inspiring and productive to work with Josef Falkinger as a co-author.

I am also more than grateful to Fabrizio Zilibotti. It has made me very happy that he agreed to be my co-supervisor. My dissertation benefited in a great deal from his useful feedback and his encouraging comments and inputs.

Further, I thank my other co-author Yingnan Zhao who has been my main collaborator at the chair. All the exchange with her and especially her mathematical skills immensely favored our works and also my own project.

I want to thank my cherished friends, in particular Simon Alder, Timo Boppart, Steve Heinke, Bea Kraus, Johannes Kunz, Igor Letina, Cornelia Rösler, Philippe Ruh, and Franziska Weiss, for making my time at the University of Zurich very pleasant and unforgettable. Furthermore, I thank Christine Le Grand and all other current and past members of the chair for all their assistance.

I also wish to express my gratitude to my parents and my two sisters. They always infinitely support and fully believe in what I do. From the deepest of my heart, I thank Christian for his never-ending understanding, encouragement and patience.

Sabrina Studer, Zurich, November 2015

Contents

I	Dissertation Overview	1
II	Research Papers	9
1	Explaining structural change towards and within the financial sector	11
1.1	Introduction	11
1.2	Model	14
1.2.1	Model set-up	14
1.2.2	Saving decision and portfolio choice	15
1.2.3	Production of goods (X-sector)	16
1.2.4	Production of financial services (Z-sectors)	17
1.3	Production equilibrium and supply of goods and financial services	18
1.3.1	Wages and prices	19
1.3.2	Resource constraints	20
1.4	Income distribution and aggregate demand	21
1.4.1	Individual saving and expenditure behavior	22
1.4.2	Aggregate demand for goods and financial services	24
1.5	The effect of the skill premium on the sectoral structure	25
1.5.1	Within change	27
1.5.2	Between change	28
1.6	General equilibrium	30
1.6.1	Existence, uniqueness and stability of equilibrium	30
1.6.2	Equilibrium skill premium	31
1.6.3	Structural change between production and financial service sectors	35
1.6.4	Structural change within the financial sector	35
1.7	Extensions	37
1.7.1	Fixed costs in the financial sector	37
1.7.2	Rents in the financial sector	38
1.7.3	Distorted portfolio choice	40
1.7.4	Participation constraints	40

1.7.5	Set-up capital for firms	42
1.7.5.1	Consumer problem	42
1.7.5.2	Firm entry and production in the X -sector	43
1.8	Robustness	45
1.9	Empirical evidence and numerical exercises	48
1.9.1	Empirics	48
1.9.1.1	Data description	48
1.9.1.2	Empirical trends	50
1.9.2	Numerics	53
1.9.2.1	Calibration	53
1.9.2.2	Numerical exercises	55
1.10	Conclusion	57
2	An equilibrium model with diversification-seeking households, competing banks and (non-)correlated financial innovations	61
2.1	Introduction	61
2.2	Model	66
2.2.1	Financial products	66
2.2.1.1	Projects with mean-variance return generating technology	66
2.2.1.2	Design of financial products and diversification effect . . .	67
2.2.2	Households' mean-variance portfolio choice	68
2.2.3	Banking sector	70
2.2.3.1	Technology structure	70
2.2.3.2	Profit function	71
2.2.3.3	Banking sector equilibrium: Competition and optimal offer	72
2.3	Equilibrium analysis	74
2.3.1	Equilibrium characterization	74
2.3.2	Effects of fundamentals on financial innovations	75
2.4	Correlated financial products and neglected correlation	76
2.4.1	Correlated derivatives	77
2.4.2	Neglected correlation	77
2.4.2.1	No neglect	77
2.4.2.2	(Partial) correlation neglect	78
2.4.3	Consequences of correlated derivatives and neglected correlation . .	80
2.5	Endogenous supply of correlated derivatives	81
2.5.1	Cost-minimal provision of perceived diversification	82
2.5.2	Equilibrium with neglected correlation	85
2.6	Intuition for immense increase in the number of financial products	86

2.6.1	Empirical features	86
2.6.2	Model intuition	87
2.7	Consequences of financial innovations	89
2.7.1	Mean-variance effects: Diversification vs. volatility	89
2.7.2	Aggregate output effects	89
2.7.3	Banking sector size	91
2.8	Conclusion	92
3	Bank lending and firm dynamics in general equilibrium	95
3.1	Introduction	95
3.2	Model	98
3.2.1	Model set-up	98
3.2.2	Households	98
3.2.2.1	Workers	99
3.2.2.2	Entrepreneurs	100
3.2.3	Financial intermediaries	102
3.2.4	Dynamic lending contract	103
3.2.4.1	Optimal financial contract	104
3.2.4.2	Theoretical properties	105
3.3	Aggregation and general equilibrium	107
3.3.1	Aggregation	108
3.3.1.1	Aggregation of workers	108
3.3.1.2	Aggregation of entrepreneurs	108
3.3.1.3	Aggregation of banks' equity	109
3.3.2	Equilibrium conditions	110
3.3.3	Definition of general equilibrium	111
3.4	Calibration and numerical results	112
3.4.1	Three optimization problems	113
3.4.1.1	Workers' optimal decisions	113
3.4.1.2	Entrepreneurs' optimal capital and labor employment	114
3.4.1.3	Banks' optimal financial contract	115
3.4.2	General equilibrium	117
3.4.3	Firm distributions	119
3.4.4	Firm dynamics	121
3.5	Discussion of dynamic programming	122
3.5.1	Starting value problems	123
3.5.2	Extrapolation errors	123
3.5.3	Sensitivity to parameter values and functional forms	124

3.5.4	Simulation issues	124
3.6	Conclusion	124

III Appendices 127

A Appendix: Chapter 1 129

A.1	Portfolio choice	129
A.2	Corner solution for securities demand	131
A.3	Further proofs	132
A.4	Data survey years 1995-2009	134

B Appendix: Chapter 2 135

B.1	Derivations	135
B.1.1	Derivation of the equilibrium	135
B.1.1.1	Derivation of N^*	135
B.1.1.2	Derivation of N_D^*	135
B.1.1.3	Derivation of $N_{D,NC}^*$	136
B.1.2	Derivation of $\sigma_B^2(N, \nu, \epsilon)$	136
B.1.3	Derivations of analytical results for endogenizing ν	136
B.1.3.1	Derivation of cost-minimizing ν^*	136
B.1.3.2	Derivation of $\underline{\epsilon}$ and $\tilde{\epsilon}$	137
B.1.3.3	Derivation of comparative statics on ν^* and N^*	138
B.1.3.4	Derivation of comparative statics on M^*	138
B.1.3.5	Derivation of comparative statics on the cost function	138
B.2	Comparative statics of fundamentals	139
B.3	Equilibria equations	139
B.3.1	Equilibria with correlated derivatives and neglected correlation	139
B.3.2	Planned and true utility levels	140
B.4	Simulation for endogenizing ν	140
B.4.1	Simulation procedure	140
B.4.2	Simulation results: Equilibrium effects of ϵ	142

C Appendix: Chapter 3 143

C.1	Timing	143
C.2	Derivations	143
C.2.1	Derivations of financial contract properties	143
C.2.1.1	Proof of Proposition 3.1	144
C.2.1.2	Proof of Lemma 3.1	144

C.2.1.3	Proof of Lemma 3.2	145
C.2.1.4	Proof of Proposition 3.2	146
C.2.1.5	Proof of Proposition 3.3	146
C.2.2	Derivation of the goods market clearing condition	148
C.3	Numerical procedure for optimization problems	149
C.3.1	Numerical procedure: Workers	149
C.3.2	Numerical procedure: Financial contract	150
C.4	Simulation procedure for general equilibrium	151
C.5	Algorithm to find general equilibrium	153
C.6	Intuition for convergence to stationary equity level	154
C.7	Additional figures	156
C.7.1	Life paths	156
C.7.2	Development of entrepreneurs' distributions	157

IV Bibliography **159**

V Curriculum Vitae **171**

Part I

Dissertation Overview

Dissertation Overview

Substantial changes in the financial sector and its relation to the real economy have characterized the last decades. Especially, the financial crisis of 2007-2008 followed by economic disruptions triggered a public debate about the financialization of the economy. These events also fostered academic research on the topic. In particular, it has been acknowledged that the interlinkage of the real economy and the financial sector is not only beneficial. This cumulative dissertation assesses the interrelation of the real economy and the financial sector by looking from a theoretical perspective at three different aspects:

First, over the last decades, there was structural change towards and within the financial sector. On one hand, the GDP share of the financial sector (finance and insurance) in the U.S. increased from below 3% in the 1940s to above 7% nowadays (data from BEA). On the other hand, the structure within the financial sector has changed. This is, “new” finance (consisting of securities, commodity contracts, investments and funds, trusts, and other financial vehicles) gained in importance compared to more “traditional” finance (consisting of Federal Reserve banks, credit intermediation and related activities and insurance carriers and related activities). In the U.S., the within finance value added share of the “new” finance subsector increased from 8% in the 1970s to around 22% in recent years (based on data from BEA). Clearly, economists want to understand what drove this two-fold structural change.

Second, another salient feature of financialization is the immense increase in the number of financial products over the last decades. For example, the number of listed securities at the Swiss Exchange SIX rose from around 3,000 in 1995 to nearly 35,000 in 2014 (data from SIX). Although exchange traded financial products only partly mirror the whole development, they clearly reflect a strong product innovation dynamic within the financial sector. It raises the questions what has driven this drastic proliferation of new financial products and what are the consequences for the economy?

Third, the interconnection between the real economy and the financial sector is also relevant at the entrepreneurial level. Access to financing is crucial for entrepreneurs; it impacts on firms’ production and shapes important characteristics of firm dynamics such as size distribution, growth and volatility of growth of firms. A crucial problem of external financing through banks loan is information asymmetry between the borrower and the lender. The questions arise how the asymmetry affects the optimal lending contracts

offered by banks to entrepreneurs and how it impacts on firms' production.

The three different aspects of the connection of the real economy and the financial sector are discussed in this dissertation in three independent chapters:

In the first chapter *Explaining structural change towards and within the financial sector* (which is joint work with Josef Falkinger and Yingnan Zhao) we analyze whether the two-fold structural change towards and within the financial sector is triggered by common drivers. We want to understand if possible drivers come from the supply side (i.e., relative price effects) or from the demand side (i.e., income effects) and how the drivers are linked to increased inequality. The questions are answered in a 3-sector / 3-factor OLG-macroeconomic model. Thereby, the baseline analysis focuses on a “perfect-world” which considers fundamentals rather than frictions as drivers. The three sectors are the goods sector, used for consumption and investment, and two finance subsectors which provide services for the transformation of savings into future consumption possibilities. The three factors are high- and low-skilled labor used in all three sectors and capital used in safe and risky return-generating technologies in the goods sector. Young households work, consume and save and old households consume savings plus returns on savings. The saving and portfolio decision of the young households is based on quasi-homothetic preferences. The demand for financial services arises from the need of transformation of savings: Financial services from the “traditional” subsector 1 are needed to transform safe savings and financial services from the “new” subsector 2 are needed to transform risky savings. In the general equilibrium the skill premium (i.e., relative wage of high- to low-skilled labor), for which demand equals supply in all three sectors and the three factor markets are cleared, is determined. Comparative-static analysis reveals channels which simultaneously drive the two-fold structural change with respect to finance and increased inequality: The main mechanism is that with increasing income people demand relatively more high-skilled services from the financial sector and especially from the finance subsector 2 (income effect). More specifically, the channels which drive the two-fold structural change and inequality simultaneously are: (i) uniform productivity progress across factors and sectors raising the wage level, (ii) skill-biased technical change that increases the skill premium and (iii) market completeness which makes risky savings more attractive. In several extensions, frictions are added to the “perfect-world” baseline model (fixed costs, rents or participation constraints in the financial sector, distorted portfolio choices or set-up capital for firms), which further impact on inequality and the sectoral structure of the economy. Finally, the theoretical model is brought to the data. The numerical exercise illustrates the increase in inequality and the two-fold structural change as observed in the U.S. over the last decades.

In the second chapter titled *An equilibrium model with diversification-seeking house-*

holds, competing banks and (non-)correlated financial innovations I focus on the drastic increase in the number of financial products. In an equilibrium model, the determinants and possible consequences of the number of financial products are analyzed. The model endogenizes the number of financial products and thereby allows distinguishing between fundamentals and (wrong) beliefs as determinants of the number of financial products. In the benchmark model, financial products are based on independent real investment projects with technological uncertainty. A financial innovation is a new risky financial product based on a new investment project. There is a diversification effect as more independent financial products are issued. Households with mean-variance preferences invest their endowment into a financial portfolio. They demand, for a fee, a bundle consisting of the risky financial products and a safe asset to generate returns for later consumption. Households are diversification-seeking in the sense that they invest relatively more in the risky bundle if there is more diversification what makes them better off. Banks supply the financial products and manage financial portfolios for a fee. They are characterized by a convex cost function and act in a competitive market. The equilibrium number of financial products and the corresponding fee are determined by competition (zero profit) in the banking sector and the investment behavior (optimality condition) of the households. Comparative-static analysis of changes in the fundamentals shows that the number of financial products increases if (i) real investment projects are more productive (higher return or lower risk), (ii) a larger volume of wealth is invested, or (iii) costs in the banking sector are lower. The benchmark model is extended to assess the impact of correlated derivatives and wrong beliefs on the number of financial products. Correlated derivatives are financial products, which are based on already existing financial products rather than on new independent real investment projects. Households have wrong beliefs if they neglect (part of the) correlation. The model predicts that correlated derivatives waste resources because they do not add to diversification and that they bias portfolio choices if their correlation is (partly) neglected. Furthermore, under correlation neglect banks deceive households by issuing more correlated derivatives instead of new financial products based on real investment projects because this lowers their costs of provision. Thus, new financial products are not always the result of changes in fundamentals. They can also be induced by investors' neglected correlation and the subsequent cost-minimizing "cheating" behavior of banks. At the normative level, these results support policy recommendations (e.g., patenting or rating of new financial products) which help to protect investors and the real economy from correlated financial products. Finally, the model allows analyzing the consequences of new financial products. It can explain a positive relationship between aggregate output, its volatility, the size of the banking sector and the number of financial products.

In the third chapter *Bank lending and firm dynamics in general equilibrium* (which is joint work with Yingnan Zhao) we look at the interrelation of the real economy and the financial sector at the entrepreneurial level. A model with dynamic long-term lending contracts between banks and firms is developed, which addresses asymmetric information in a general equilibrium framework. There are ex-ante identical, risk-averse households who decide to become either worker or entrepreneur. Workers supply labor, consume and save. Entrepreneurs run firms and produce by employing labor and capital. Their production function exhibits decreasing returns to scale; the productivity realization is state-contingent and private information to the entrepreneurs. There are risk-neutral banks that offer annuity deposits to workers and use workers' savings plus own bank equity to finance firms' production. To resolve the information asymmetry (banks cannot observe the realized state of productivity), banks offer to entrepreneurs profit-maximizing long-term dynamic lending contracts with bank loans and repayments. The optimal contracts are determined by recursive dynamic programming with promised value as state variable. The promised value is the continuation utility of an entrepreneur from its future consumption of net revenue (output generated by using bank loans minus repayments). Optimal contracts are promise keeping, incentive compatible and fulfill the limited liability and the credibility constraints. The general equilibrium model is analyzed numerically. First, the three partial parts are solved: (i) Workers consume and save more, but supply less labor, if they hold more current period deposits. (ii) Entrepreneurs employ a capital intensity for production independent of the level of bank loans they receive. (iii) The recursive optimal contract is characterized as follows: The higher today's promised value the higher are the level of bank loans and state-contingent future promised values. State-contingent repayments first increase in the level of promised values and then decrease. The contract induces entrepreneurs' truth-telling about productivity realizations by postponing reward for a high productivity and by postponing punishment for a low productivity. Second, the three partial parts are combined to determine the general equilibrium: By simulating the economy with many households of different ages (for the entrepreneurs this includes different histories of productivity realizations), the equilibrium interest and wage rates as well as the share of entrepreneurs are determined by clearing in the labor and the capital market and by zero profits for banks. In equilibrium, the interest rate is around 4% and the share of entrepreneurs is around 8%; numbers are in line with literature and empirics. Furthermore, the simulation of different paths of life of entrepreneurs allows determining the distributions of firms in the economy and firm dynamics: On average, older firms are larger and grow less but in a more stable way than younger firms. In addition to numerically solving the model and providing economic explanations for the results, computational issues of the considered in dynamic programming problem are discussed.

This dissertation adds to a better understanding of the interrelation of the real economy and the financial sector by connecting them in different theoretical general equilibrium models. It is worked out how fundamentals – both from the real economy and the financial sector – as well as (distorted) beliefs and behaviors – of both households and banks – lead to changes in the financial sector and its relation to the real economy. In particular, it is shown how the interactions of agents from the real economy and from the financial sector affect the sectoral structure of the economy; how they determine the number of financial products; and how producing firms with access to bank loan financing evolve, respectively. The structure of the dissertation is as follows: The three papers are found in Part II and the respective appendices are provided in Part III. Part IV contains the bibliography and Part V presents my curriculum vitae.

Part II

Research Papers

1 Explaining structural change towards and within the financial sector¹

Joint with Josef Falkinger and Yingnan Zhao

1.1 Introduction

Financialization and inequality are topics that stir up the public debate – among experts as well as outside the scientific community. Discussions about financialization (Greenwood and Scharfstein, 2013; Philippon and Reshef, 2012, 2013) have gained momentum by the financial crisis; the inequality debate was brought “in from the cold” (Atkinson, 1997) towards the end of the last century and has reached the center court recently with the Piketty book (Piketty, 2014). This paper argues that the two phenomena are genuinely related to each other. Structural change towards and within the financial sector, as observed over the last three decades, enhances inequality. And rising inequality fosters financialization.

We present our argument in a model that comprises the most basic tools provided by economics for analyzing sectoral structure and distribution. Financialization means two things: The weight of financial business relative to non-financial business increases and the type of financial business changes. From a macroeconomic perspective the first aspect can be summarized as structural change towards the financial sector: The financial sector expands relative to the production sector. We do not approach this question from a monetary or financial aspect like the nominal transaction volume of the financial relative to the real sector. Our perspective is a real economics one: The financial sector employs resources and generates income for the resources employed. That is, there must be some kind of output (service) that is produced, sold and purchased. The relevant measures are therefore employment and income or output shares; the essential components to be modeled are the production function of the financial sector and the demand function for financial services. For capturing the second aspect of financialization – the shift from conventional banking type activities to sophisticated modern finance – an appropriate

¹This chapter is a revised version of the Working Paper No. 206 from the Working paper series / Department of Economics at the University of Zurich.

model structure requires two separate subsectors within the financial sector which differ in their demand and production characteristics. In sum, we have therefore a three-sector model – one production sector and two financial subsectors.

Inequality requires incorporating heterogeneous agents which differ in their endowments. In our model we have low-skilled and high-skilled workers. They are mobile between sectors and cost-minimal skill-intensities differ across sectors. As a consequence, the interaction between sectoral structure and inequality comes through the skill premium. The focus on inequality between low-skilled and high-skilled workers is on the one side motivated by the empirical fact that the rise in inequality over the last decades has been driven to a large extent by skill premia and skill composition, as the ample evidence from the skill-bias literature shows (for instance, Machin and Van Reenen (1998); Piketty and Saez (2003)). On the other side, we see it as a first important step, which later might be complemented by elements which focus on the functional distribution of income between workers and capitalists or on rents. There is capital in our model; it must be. After all, financial markets have the purpose to transform, under risk, current resources into future production possibilities. This requires, on the one side, saving decisions and, on the other side, capital investment into revenue-bearing inputs to future production. In our model, returns on capital are generated by two different types of technologies (robust and risky) which transform savings into future consumption possibilities.

Structural change can be caused by the supply side: Changing endowments or technical change. The huge literature on directed technical change, for instance, has emphasized this channel (Acemoglu, 2002). There is, however, also an important role for the demand side. Although often neglected, income effects are essential for aggregate developments (Boppart, 2014, 2015; Föllmi and Zweimüller, 2008). We account for demand side effects by assuming that agents have quasi-homothetic preferences of the Stone-Geary form. The specific finance aspect enters the demand side of our model through the following channel: Demand for financial services comes from the need to manage portfolios and to finance investments into profitable projects in a way that reflects the preferences of the agents who own the endowments of the economy. Stone-Geary preferences account for the fact that a part of the savings is motivated by future subsistence expenditures.

In our model the finance industry correctly assesses risks and productivity of investment projects and earns no rents. This is against popular views; neither does it reflect the view of the authors of this paper. Actually, there are many sources for imperfections in the financial sector. For instance, financial products may be distorted by neglected correlation (see Chapter 2 of this dissertation), or insider knowledge and barriers to entry generate rents for financial intermediation. A salient example is the so-called finance premium. There is convincing evidence that a finance premium exists (Célérier and Vallée,

2015; Philippon and Reshef, 2007, 2012), that is, the same type of labor earns more in a finance job than in other occupations. Nonetheless, from a methodological point of view, we consider it as important to start with a benchmark model in which distortions are kept at a minimum. Given the firm basis of such a benchmark, one can then be bold in looking at the consequences of imperfections which certainly exists in reality in general and in the financial business in particular. Arguably, rents can be more easily extracted when they go along with the tide rather than against it. So it is important to know if outcome changes are supported by changes in economic fundamentals. Section 1.7 gives extensions which provide some ideas how distortions affect the comparative-static results of the benchmark model. Moreover, in the quantitative implementation of our model in Section 1.9, we try to separate the rent component of the expansion of the financial sector, in particular new finance, from the part that is driven by economic fundamentals.

There is a long literature on the impact of financial development on economic growth (Levine, 2005).² The causes of financial sector growth and the changing structure of financial activities, which are the topic of this paper, have been less scrutinized. The literature related to our paper in a more narrow sense is rich as far as the empirical side is concerned. In particular, Philippon and his co-authors did pioneering empirical work on financialization. On the theoretical side the situation is quite different. To our knowledge there are only two attempts to explain structural change towards finance in a general equilibrium framework. Philippon (2012) sketches in his notes a 2x2 model with a real and a financial sector both producing with capital and labor. The financial sector produces intermediation services for households and firms. The focus is on the equilibrium effects of changes in intermediation costs. Improvements in financial intermediation tend to raise real wages but have in general an ambiguous effect on the GDP-share of the financial sector. The GDP-share of finance rises if more firms need intermediation services. Structural change between services for safe assets and services for risky investments or wage inequality are not addressed nor do income effects play a role for the relative size of the financial compared to the real sector. There is only one type of labor, one interest-bearing asset and preferences are homothetic. Moreover, there are two types of households - infinitely living saver households and households which live two periods and borrow when young. By contrast, in our paper all households live for two periods and save when young;

²While the dominant view in this literature was that financial development is positive for growth, a more skeptical view has emerged in the recent past. Gründler and Weitzel (2012) or Law and Singh (2014) provide evidence that more finance is good for growth at low levels of financial development but harmful beyond a certain threshold. Financial sector growth seems to harm in particular skill-intensive (Kneer, 2013) and R&D intensive (Cecchetti and Kharroubi, 2015) industries. Moreover, negative growth effects are robust if different measures of financialization are used, for instance market capitalization rather than credits (Rousseau and Wachtel, 2011) or the employment share of the financial sector (Capelle-Blancard and Labonne, 2011). Beck et al. (2012) find that in particular the shift from enterprise credits to household credits is detrimental for growth and inequality enhancing.

savings can be invested in a portfolio of safe and risky assets. The second theoretical explanation of structural change towards finance is provided by Gennaioli et al. (2014). Like in Philippon (2012) a 2x2-framework is considered and structural change within the financial is not in the focus of the paper. The real sector produces with capital and labor, the financial sector consists of financial intermediation experts in whom investors trust. Therefore they are willing to pay them fees. Like in our set-up households live two periods and save when young. Moreover, they also account for risky assets. Inequality among households, however, plays no role. Their saving decision is exogenous - young households save the entire wage - and the portfolio choice is determined by mean-variance preferences. The main driver for structural change towards finance in their model is the idea that financial intermediation services are not only required for the financing of new capital but also for the preservation of the entire stock of capital accumulated over time. Since in a Solow type growth model the capital coefficient increases, the share of financial services in GDP increases, too. In our model, which focuses on comparative-static equilibrium effects of skills and endowments, technologies and preferences, no long-run accumulation effect is considered.

The structure of the paper is as follows. The next section outlines the formal structure of our 3x3-model and its building blocks. Section 1.3 analyzes the production equilibrium. Section 1.4 derives the demand for goods and financial services. Section 1.5 summarizes the effects of inequality on the sectoral structure of the economy. In Section 1.6 the general equilibrium is characterized and comparative-static effects are derived analytically. Section 1.7 gives extensions which provide some ideas how distortions affect the comparative-static results of the benchmark model. In Section 1.8 an alternative to the benchmark specification of the model is considered and the robustness of the results is discussed. Section 1.9 confronts the theoretical results with empirical evidence from the U.S.. Moreover, a numerical exercise is provided. Main conclusions are summarized in Section 1.10.

1.2 Model

1.2.1 Model set-up

We model a 3-sector, 3-factor economy. There is a production sector X and a finance sector Z with two subsectors Z_1 and Z_2 . All sectors employ low-skilled and high-skilled workers. Produced goods are used for consumption and investment. For transforming savings into future consumption possibilities, more or less risky technologies are available which use capital as input and deliver consumption goods as output in the next period.

(As an extension we present a variant of the model, in which capital is used in the X sector to set up producing firms.) Financial services have the function to support the transformation of savings into future consumption possibilities. Services Z_1 are used for safe savings. Services Z_2 provide state-dependent instruments and are used for savings in securities with risky returns.

We consider a (static) two-period OLG economy. The future $t = 1$ is uncertain. It consists of a set Θ of distinguishable events and a set $\bar{\Theta}$ of events which are indistinguishable in $t = 0$. The future state space is $\{\{\theta|\theta \in \Theta\}, \bar{\Theta}\}$. We have $\text{prob}(\Theta)=\mu$ and $\text{prob}(\theta|\Theta)=\pi_\theta$ with $\sum_{\theta \in \Theta} \pi_\theta = 1$.³ For $\theta \in \Theta$, state-contingent investment possibilities are available which pay off if and only if state θ is realized. No state-contingent investment possibilities exist for $\bar{\Theta}$ which reflects “true uncertainty”.

1.2.2 Saving decision and portfolio choice

There are N agents who live for two periods. They are endowed with a skill level and work as either high-skilled or low-skilled worker when young. The number of low-skilled workers is \bar{L} and the number of high-skilled workers is \bar{H} . The efficiency units of labor provided by a high-skilled and a low-skilled agent are given by b_H and b_L , respectively. They are paid a wage per efficiency unit at rate, $w_l, l \in \{L, H\}$. Income $y^l = w_l b_l$ can be consumed in $t = 0$ or be saved and transformed to tomorrow’s consumption possibilities. Agents are assumed to have quasi-homothetic preferences of the Stone-Geary form: Beyond a subsistence level to be expended they spend income on the good produced in the X -sector.⁴ They have an instantaneous indirect utility function of the form $\log(e_t - \bar{e}_t)$ where e_t is the expenditure for good X consumption and $\bar{e}_t \geq 0$ is the subsistence expenditure level in time t . Intertemporal preferences are assumed to be additive logarithmic with a discount factor δ .

The intertemporal problem of agents consists of two parts: A saving decision and a portfolio choice. On the one hand, agents have to decide how much to expend on consumption, e_0 , and how much to save, s . On the other hand, they have to put the saving in an appropriate portfolio of financial products. For this purpose they demand financial services. With the support of these services they decide how much of the saving is put into deposits, d , with a safe payoff r , and how much into risky state-contingent financial products (Arrow securities), f_θ , which pay off R_θ if state θ is realized and zero otherwise. We assume that all Arrow securities have the same expected payoff. Specifically, there

³This structure is taken from Falkinger (2014).

⁴Achury et al. (2012) show that a Stone-Geary type utility function is appropriate for explaining stylized facts of household finance like higher saving rates of households with higher lifetime income or a larger fraction of risky assets in the portfolios of wealthy agents.

exists $R > 0$ so that

$$R_\theta = \frac{R}{\pi_\theta}, \quad \theta \in \Theta. \quad (1.1)$$

For transforming one unit of deposit, one unit of financial services from subsector 1 is needed; and for transforming one unit of Arrow securities, one unit of financial services from subsector 2 is required. Therefore, given the portfolio choice, $\{d, f\}$, with $f = \sum_{\theta \in \Theta} f_\theta$, agents have to pay a fee $T = p_{z_1}d + p_{z_2}f$ to the financial sector, where p_{z_1} and p_{z_2} are the prices for financial services Z_1 and Z_2 , respectively.⁵ Suppose the fee is charged in the first period and agents internalize the fee in their portfolio choice. The expected utility maximization problem of an agent l with income y^l is then given by:

$$\max_{s^l, \{f_\theta^l\}_{\theta \in \Theta}, d^l} \mathbb{E}U = \log(e_0^l - \bar{e}_0) + \delta \left[\mu \sum_{\theta \in \Theta} \pi_\theta \log(e_\theta^l - \bar{e}_1) + (1 - \mu) \log(e_\Theta^l - \bar{e}_1) \right]$$

subject to

$$e_0^l + (1 + p_{z_1})d^l + (1 + p_{z_2}) \sum_{\theta \in \Theta} f_\theta^l = y^l \quad (1.2)$$

$$e_\theta^l = \begin{cases} R_\theta f_\theta^l + r d^l, & \text{if } \theta \in \Theta \\ r d^l, & \text{otherwise} \end{cases} \quad (1.3)$$

$$s^l = \sum_{\theta \in \Theta} f_\theta^l + d^l. \quad (1.4)$$

In Section 1.4 aggregate demand functions for goods and financial services are derived from this program.

1.2.3 Production of goods (X-sector)

Firms in the X -sector employ low-skilled and high-skilled labor as input factors in a linear homogeneous production function

$$X = G^x(H_X, L_X),$$

where H_X, L_X denote respective labor employment in the X -sector. There is perfect competition with zero-profit prices. This means:

$$p_x = c_x(w_H, w_L), \quad (1.5)$$

⁵Without loss of generality, it was assumed that financial services are measured in units of savings. Without this normalization the cost of financial services per unit of saving would be $\tilde{p}_{z_i} = p_{z_i}n_i$ rather than p_{z_i} , where n_i denotes the units of financial services needed for one unit of saving in deposits ($i = 1$) and securities ($i = 2$), respectively.

where $c_x(w_H, w_L)$ are the unit costs and w_H, w_L are the wage rates of high- and low-skilled labor, respectively, per efficiency units.

The goods price is taken as numéraire, $p_x = 1$. Revenue X is distributed to labor as follows:

$$W_x = w_L L_x + w_H H_x = G^x(L_x, H_x),$$

where W_x is total wage earned in the X -sector.

Capital is used in technologies which transform savings into future consumption possibilities. Two types of technologies are available: A robust technology, which transforms under any condition (i.e., in Θ and $\bar{\Theta}$) one unit of capital invested today into r units of output tomorrow; furthermore, for $\theta \in \Theta$, a set of risky technologies specialized to θ -contingent environments. One unit of capital invested in technology θ delivers R_θ units of output if state $\theta \in \Theta$ occurs tomorrow and zero otherwise. Deposits are invested in the robust technology; savings in securities are invested in the respective risky technologies. The smaller the measure π_θ of the state to which a risky technology is targeted, the more productive the capital invested in the technology. (1.1) expresses this relationship between specialization advantage and risk.

The separation of the production of old age consumption goods by capital from the labor-based production of the goods consumed and invested in the active period of life is convenient from an analytical point of view. Under a more realistic perspective, however, capital is typically a prerequisite for producing with labor. In the extension in Section 1.7.5, we show that essentially the same payoff structure arises if X is produced under monopolistic competition and capital is needed to set up firms – by robust and risky set-up technologies, respectively. Asset returns are then generated by the operating profits of the firms the set-up of which has been financed by the asset.

In almost all of the further analysis only the relative payoff between robust and specialized risky technologies matters. It is given by:

$$\rho \equiv \frac{r}{R}.$$

The only exception is the discounting of future subsistence expenditure, $\frac{\bar{e}_1}{r}$, for which the level of the return on the robust technology will matter.

1.2.4 Production of financial services (Z -sectors)

The financial sector Z consists of two subsectors, Z_1 and Z_2 . They provide financial services for transforming savings through safe and risky assets into future consumption possibilities. (The assets are invested in the robust and risky technologies and households get the generated revenue as return on their investment.)

Z_i , $i \in \{1, 2\}$, is produced with a linear homogeneous production function $G^{z_i}(\cdot)$:

$$Z_i = G^{z_i}(H_{z_i}, L_{z_i}), \quad i \in \{1, 2\} \quad (1.6)$$

where H_{z_i} , L_{z_i} denote employment levels in the Z_i -sector.

In reality, fixed costs may play an important role in the provision of financial services. We consider such costs as an extension in Section 1.7.1 and show how changes in fixed costs affect the equilibrium outcomes of our model.

We assume perfect competition in the Z -sectors and have therefore zero-profit prices

$$p_{z_i} = c_{z_i}(w_H, w_L), \quad i \in \{1, 2\} \quad (1.7)$$

where $c_{z_i}(w_H, w_L)$ are the unit costs.

Revenue $p_{z_i}Z_i$, $i \in \{1, 2\}$, is distributed to labor

$$W_{z_i} = w_L L_{z_i} + w_H H_{z_i} = p_{z_i} G^{z_i}(H_{z_i}, L_{z_i}), \quad i \in \{1, 2\}$$

where W_{z_i} is total labor income earned in the Z_i -sector.

As emphasized in the introduction of this paper, perfect competition in the Z -sector is an ideal benchmark rather than a description of reality. The role of rents is considered in the extension presented in Section 1.7.2.

1.3 Production equilibrium and supply of goods and financial services

At the production side, the essential feature we want to address is variation in skill intensities. For an explicit comparative-static analysis we take production functions of the Cobb-Douglas form.

Let, for $j \in \{x, z_1, z_2\}$, G^j have Cobb-Douglas form

$$G^j(L_j, H_j) = A_j L_j^{1-\alpha_j} H_j^{\alpha_j},$$

where A_j is total factor productivity and α_j is the factor share of high-skilled workers in sector j .⁶ Then

$$a_j^L = \frac{1}{A_j \kappa_j^{\alpha_j}}, \quad a_j^H = \frac{\kappa_j^{1-\alpha_j}}{A_j} \quad (1.8)$$

⁶The magnitudes of the total factor productivities depend on the unit in which financial services are measured. Since financial services are measured in units of savings, $A_x < A_{z_1} \leq A_{z_2}$ is a plausible restriction on total factor productivities. Analytically no such restriction is required for the results.

are the input coefficients, and the cost-minimizing skill-intensities $\kappa_j \equiv a_j^H/a_j^L$ are given by

$$\kappa_j(\omega) = \frac{\gamma_j}{\omega}, \quad \gamma_j \equiv \frac{\alpha_j}{1 - \alpha_j}, \quad (1.9)$$

where $\omega \equiv w_H/w_L$ is the relative wage per efficiency unit of skilled labor compared to unskilled labor. It reflects the skill premium (per efficiency unit).⁷

1.3.1 Wages and prices

We have for variable unit costs in sector j :

$$c_j(w_H, w_L) = \frac{w_L^{1-\alpha_j} w_H^{\alpha_j}}{A_j \Gamma_j}, \quad \Gamma_j \equiv \alpha_j^{\alpha_j} (1 - \alpha_j)^{1-\alpha_j}. \quad (1.10)$$

Using (1.10) and $p_x = 1$ in the zero-profit price equation (1.5), we obtain

$$w_L = A_x \Gamma_x \omega^{-\alpha_x}, \quad (1.11)$$

and from (1.7), for $i \in \{1, 2\}$,

$$p_{z_i} = \frac{A_x}{A_{z_i}} \frac{\Gamma_x}{\Gamma_{z_i}} \omega^{\alpha_{z_i} - \alpha_x}. \quad (1.12)$$

In sum, prices for financial services are related to the skill premium in the following way:

Fact 1.1. *The price of financial services Z_i , p_{z_i} , is an increasing function of ω if $\alpha_{z_i} > \alpha_x$. If $\alpha_{z_i} = \alpha_x$, then p_{z_i} is invariant with respect to ω . Moreover, $\alpha_{z_i} > \alpha_x$ ($\alpha_{z_i} = \alpha_x$) is equivalent to $\kappa_{z_i} > \kappa_x$ ($\kappa_{z_i} = \kappa_x$).*

Proof. Follows from (1.12) and (1.9), respectively. \square

As known from the Stolper-Samuelson theorem, this fact holds quite generally and is not an artifact of the Cobb-Douglas specification.

In the further analysis we make the following assumption about the factor intensity ranking of the three sectors:

Assumption 1.1. $\alpha_{z_2} \geq \alpha_{z_1}$ and $\alpha_{z_1} \geq \alpha_x$ with at least one inequality holding strictly.

In Section 1.9 we provide evidence on the sectoral skill intensities. Assumption 1.1 is consistent with the evidence.

⁷Note that $\kappa_j = \frac{b_H \bar{H}_j}{b_L \bar{L}_j}$. According to (1.9), the inverse labor demand function is $\omega = \left(\gamma_j \frac{b_L}{b_H} \right) \frac{\bar{L}_j}{\bar{H}_j}$. Thus, we have skill-biased technical change (in the sense of an outward shift of skilled-labor demand relative to unskilled-labor demand) if the output elasticity α_j of high-skilled labor rises or if there is low-skilled labor augmenting progress (that is b_L/b_H rises). It is worth noting that α_j is a sector-specific component whereas b_L/b_H is uniform across sectors.

1.3.2 Resource constraints

Total labor endowment in efficiency units is given by

$$L = b_L \bar{L}, \quad H = b_H \bar{H},$$

so that the “skill richness” of the total labor force is

$$k \equiv \frac{b_H \bar{H}}{b_L \bar{L}}.$$

The aggregate resource constraints are:

$$\begin{aligned} a_x^L X + a_{z_1}^L Z_1 + a_{z_2}^L Z_2 &= b_L \bar{L} \\ a_x^H X + a_{z_1}^H Z_1 + a_{z_2}^H Z_2 &= b_H \bar{H} \end{aligned} \tag{1.13}$$

with a_j^l , $j \in \{x, z_1, z_2\}$, $l \in \{H, L\}$ being functions of the skill premium ω defined in (1.9).

For illuminating the drivers of structural change on the production side it is worth looking, as an intermediary step, separately at the allocation of resources within the financial sector and the resource allocation between financial services and goods production.

We focus first on the allocation within the financial sector. Let total employment (in efficiency units) in the financial sector be given by L_z and H_z , respectively. If $\alpha_{z_2} = \alpha_{z_1}$, the allocation of L_z and H_z on Z_1 and Z_2 is determined by the demand side only. If $\alpha_{z_2} > \alpha_{z_1}$, then the resource constraints $a_{z_1}^L Z_1 + a_{z_2}^L Z_2 = L_z$ and $a_{z_1}^H Z_1 + a_{z_2}^H Z_2 = H_z$ solve to:

$$Z_1 = \frac{L_z(\kappa_{z_2} - k_z)}{a_{z_1}^L(\kappa_{z_2} - \kappa_{z_1})}, \quad Z_2 = \frac{L_z(k_z - \kappa_{z_1})}{a_{z_2}^L(\kappa_{z_2} - \kappa_{z_1})}, \tag{1.14}$$

where $k_z \equiv \frac{H_z}{L_z}$ is the “skill richness” of the labor force in the financial sector. This implies for the supply structure within the financial sector:

$$\frac{Z_2}{Z_1} = \frac{a_{z_1}^L k_z - \kappa_{z_1}}{a_{z_2}^L \kappa_{z_2} - k_z} \equiv \chi(\omega, k_z) \tag{1.15}$$

The following result on within sector structural change follows immediately.

Proposition 1.1. *If $\alpha_{z_2} > \alpha_{z_1}$, for a given level of employment in the financial sector, an increase in the skill premium or a rise in the skill richness of labor employed in the financial sector shift the supply structure from traditional financial services Z_1 to new financial services Z_2 .*

Proof. According to (1.9), $\kappa_{z2} > \kappa_{z1}$ if $\alpha_{z2} > \alpha_{z1}$. For $\kappa_{z2} > \kappa_{z1}$, $\frac{\partial \chi}{\partial \omega} > 0$ and $\frac{\partial \chi}{\partial k_z} > 0$ as known from the Rybczynski analysis. \square

Moreover, for a given level of the skill richness, k_z , of labor employed in the financial sector, system (1.13) can be written in the form

$$\begin{aligned} a_x^L X + L_z &= b_L \bar{L} \\ a_x^H X + k_z L_z &= b_H \bar{H} \end{aligned} \quad (1.16)$$

which leads to the following result.

Fact 1.2. *For a given level of skill richness in the financial sector, we have*

$$\frac{L_z}{L_x} = \frac{k - \kappa_x}{k_z - k}. \quad (1.17)$$

Proof. System (1.16) solves to $L_x = b_L \bar{L} \frac{k_z - k}{k_z - \kappa_x}$, $L_z = b_L \bar{L} \frac{k - \kappa_x}{k_z - \kappa_x}$. Assumption 1.1 implies $k_z > k > \kappa_x$. \square

Thus, for a given skill premium ω (so that κ_x is fixed) and a given skill richness k_z in the financial sector, employment in the financial sector is ceteris paribus higher in an economy with a large share of skilled labor k .

In a general equilibrium, however, employment in the financial sector is determined simultaneously with the allocation of resources to the goods sector.

1.4 Income distribution and aggregate demand

The demand for financial services comes from the need of agents to transform current savings into future income. For this purpose the asset-holding agents require financial products and expert services from the financial sector which support them by choosing and managing a portfolio of deposits and securities appropriate for their preferences.

The program $\max EU$ subject to (1.2)-(1.4) is only well-defined if $e_0 > \bar{e}_0$ and $e_1 > \bar{e}_1$. This requires that

$$y^l = b_l w_l > \bar{y} \equiv \bar{e}_0 + (1 + p_{z1}) \frac{\bar{e}_1}{r}, \quad l \in \{L, H\}. \quad (1.18)$$

\bar{y} denotes the present value of future subsistence expenditure in units of today's final output.

Assuming $y^H \geq y^L$, which is equivalent to $\omega \geq b_L/b_H$, $y^L > \bar{y}$ is sufficient for (1.18). The following fact gives a necessary and sufficient condition for $y^L > \bar{y}$. The signs below the parameters show the sign of the respective partial derivatives.

Fact 1.3. *There exists a threshold ω_L^+ so that $y^L > \bar{y}$ if and only if $\omega < \omega_L^+(A_{x,+}, A_{z_1,+}, b_{L,+}, \bar{e}_0, \frac{\bar{e}_1}{r})$.*

Proof. Appendix A.3. □

Savings in securities are positive if and only if the following condition holds: $\mu R(1 + p_{z_1}) > (1 + p_{z_2})r$. The condition can be rewritten in the form

$$\mu > p\rho, \quad p \equiv \frac{1 + p_{z_2}}{1 + p_{z_1}}, \quad \rho \equiv \frac{r}{R}. \quad (1.19)$$

$p\rho$ is the relative net payoff (i.e., after correction for costs of financial services) of savings in safe assets compared to savings in risky assets. If condition (1.19) is violated, the expected net payoff of risky investment is lower than the net payoff of risk-free investments and all savings is in deposits.

In the next subsection we analyze individual saving and expenditure behavior. Subsection 1.4.2 deals with aggregate demand.

1.4.1 Individual saving and expenditure behavior

As is derived in Appendix A.1, under the assumption that inequalities (1.18) and (1.19) are satisfied, individual savings in deposits and securities are given by

$$d^l = s_d \frac{\delta}{1 + \delta} \frac{y^l - \bar{y}}{1 + p_{z_1}} + \frac{\bar{e}_1}{r}, \quad l = \{L, H\}, \quad (1.20)$$

and

$$f^l = s_f \frac{\delta}{1 + \delta} \frac{y^l - \bar{y}}{1 + p_{z_2}}, \quad f_\theta^l = \pi_\theta f^l, \quad \theta \in \Theta, \quad l = \{L, H\}, \quad (1.21)$$

respectively, with

$$s_d = \frac{1 - \mu}{1 - p\rho}, \quad s_f = \frac{\mu - p\rho}{1 - p\rho}. \quad (1.22)$$

Apart from the savings for future subsistence expenditure, $\frac{\bar{e}_1}{r}$, in form of deposits, the saving level is proportional to the supernumerary budget $y^l - \bar{y}$. In real terms, the value of the supernumerary budget, which is relevant as a basis for saving, depends on the price of the financial service charged on the particular form of savings – p_{z_1} for deposits and p_{z_2} for securities. The split of the saving on safe and risky assets is given by the marginal propensities to save in deposits, s_d , and in securities, s_f , respectively. The propensity of safe investment increases in the relative net payoff of the safe asset, $p\rho$, and declines with the measure μ of states covered by securities. The propensity of risky investment reacts in

the opposite direction.⁸ In sum, the two propensities add up to one so that total saving, $s^l = d^l + f^l$, is given by:

$$s^l = \frac{\delta}{1+\delta} \frac{y^l - \bar{y}}{1+p_{z_1}} \left(s_d + \frac{s_f}{p} \right) + \frac{\bar{e}_1}{r} \quad (1.23)$$

If saving in securities is more costly than saving in deposits, s_f is discounted by the fee differential p .⁹

In contrast to net savings, gross savings include the fee to be paid for the financial services consumed in support for the transformation of savings into future income. Adding up $(1+p_{z_1})d^l + (1+p_{z_2})f^l$, we have

$$s^l + t^l = \frac{\delta}{1+\delta} (y^l - \bar{y}) + \frac{(1+p_{z_1})\bar{e}_1}{r}, \quad (1.24)$$

where $t^l = p_{z_1}d^l + p_{z_2}f^l$ denotes the total fee paid by agent l .

Current expenditures $e_0^l = y^l - (s^l + t^l)$ are thus:

$$e_0^l = \frac{1}{1+\delta} (y^l - \bar{y}) + \bar{e}_0. \quad (1.25)$$

For the discussion of structural change on the demand side, the effect of income on the portfolio structure is of particular importance.¹⁰ According to (1.20) and (1.21), richer agents invest a larger share of their saving in risky assets than the relatively poorer ones. The reason is that the provision for future subsistence expenditure by safe investments has diminishing weight if people become richer. This means that saving in deposits has the character of a “necessity” and saving in risky securities is a “luxury”. Moreover, if present subsistence expenditure is more pressing than future subsistence expenditure, people save a smaller part of their income when they are poor and the saving rate s/y rises when they get richer.¹¹ The following fact summarizes this important implication of our model.

⁸For $\bar{e}_0 = \bar{e}_1 = 0$ and $p_{z_1} = p_{z_2} = 0$, we have $s_d = \frac{1-\mu}{1-\rho}$ and $s_f = \frac{\mu-\rho}{1-\rho}$. Defining $\bar{R} = \frac{R}{\mu}$ and $\bar{\rho} = \frac{r}{R}$, we can rewrite the two terms in the form $s_d = \frac{\bar{R}(1-\mu)}{R-r/\mu}$ and $s_f = \frac{\mu\bar{R}-r/\mu}{R-r/\mu}$. Thus, with Cobb-Douglas preferences and zero financial intermediation cost, the portfolio choice coincides with the one in Acemoglu and Zilibotti (1997), where the conditional expectation \bar{R} of the productivity of risky technologies is used rather than the unconditional expectation R .

⁹If inequality (1.19) is violated, then saving in securities is unattractive in the first place and we have a corner solution with $s_f = 0$ and $s_d = s = \frac{\delta}{1+\delta} \frac{y-\bar{y}}{1+p_{z_1}} + \frac{\bar{e}_1}{r}$.

¹⁰Boppart (2015) analyzes the skill-content of the consumption basket of different income groups. With rising income, a household's demand shifts towards skill-intensive sectors (including financial services; also shown by Suellow (2015) in detail).

¹¹The role of subsistence requirements for the saving behavior may call into mind the effects of fixed costs in the model of Greenwood and Jovanovic (1990), where saving rate and portfolio structure depend on an agent's wealth due to constrained participation in the use of financial intermediation service. While we consider the effect of a participation constraint as an extension in Section 1.7, no such constraint exists

Fact 1.4. *Let $\bar{e}_0 > 0$ or $\bar{e}_1 > 0$.*

a) *If $\bar{e}_1 > 0$, then $\frac{\partial(f/d)}{\partial y} > 0$.*

b) *For $\bar{e}_0 > 0$, $\frac{\partial(s/y)}{\partial y} > 0$ if and only if $\frac{\delta \bar{e}_0}{1+p_{z1}} > \frac{\bar{e}_1}{r} \left[\frac{1+\delta}{s_d+s_f/p} - \delta \right]$. (Note that for $p = 1$ the square-bracketed term reduces to one.)*

Proof. Part a) follows immediately from (1.20) and (1.21). For b) the definition of \bar{y} in (1.18) is used. \square

1.4.2 Aggregate demand for goods and financial services

Saving and expenditure behavior follow affine-linear functions. Therefore, aggregate behavior depends on two things: The level of aggregate income and the number of people over which the income is distributed. The latter comes in through the fact that subsistence requirements are bound to the existence of an agent, independent of her or his income.

Aggregating the two pools of agents, we have

$$N = \bar{L} + \bar{H}$$

for the size of the population and

$$W = w_L b_L \bar{L} + w_H b_H \bar{H}$$

for the level of aggregate income. In view of (1.11), the latter amounts to

$$W = A_x \Gamma_x b_L \bar{L} \omega^{-\alpha_x} (1 + \omega k). \quad (1.26)$$

The following fact shows that aggregate income, measured in units of X , is an increasing function of the skill premium ($\omega = w_H/w_L$).

Fact 1.5. *Under Assumption 1.1, W is increasing in ω . We have*

$$\frac{\partial W}{\partial \omega} = A_w \omega^{-\alpha_x} (1 - \alpha_x) (k - \kappa_x) > 0 \quad (1.27)$$

with $A_w \equiv A_x \Gamma_x b_L \bar{L}$.

in the baseline considered here. But everybody has to expend a certain sum to survive. This biases saving rate and portfolio structure. If people get richer the pressure of the subsistence requirements diminishes. There are of course other important differences to Greenwood and Jovanovic. In particular, all forms of saving require costly financial intermediation in our framework. Moreover, our focus is on inequality in labor income rather than wealth inequality and on structural change rather than growth.

Proof. According to (1.26),

$$\begin{aligned}\frac{\partial W}{\partial \omega} &= A_w \omega^{-\alpha_x} \left[-\frac{\alpha_x}{\omega} (1 + \omega k) + k \right] \\ &= A_w \omega^{-\alpha_x} \left[-\frac{\alpha_x}{\omega} + (1 - \alpha_x) k \right] = A_w \omega^{-\alpha_x} (1 - \alpha_x) \left[k - \frac{\alpha_x}{1 - \alpha_x} \frac{w_L}{w_H} \right].\end{aligned}$$

According to (1.9),

$$\frac{\alpha_x}{1 - \alpha_x} = \frac{w_H a_x^H}{w_L a_x^L}.$$

Thus, the square-bracketed term reduces to $k - \kappa_x$, which is positive if Assumption 1.1 holds. \square

Financial services provision is more skill intensive than goods production, at least on average. Therefore, in terms of goods, aggregate wage income rises with the skill premium. A different matter is the impact of the skill premium on the purchasing power for financial services, the price of which rises too with the skill premium.

Aggregating individual investments in deposits, given by (1.20), we obtain

$$D = \left(s_d \frac{\delta}{1 + \delta} \frac{\bar{w} - \bar{y}}{1 + p_{z_1}} + \frac{\bar{e}_1}{r} \right) N, \quad (1.28)$$

where $\bar{w} \equiv \frac{W}{N}$ denotes average income. In an analogous way, we have from (1.21):

$$F = s_f \frac{\delta}{1 + \delta} \frac{\bar{w} - \bar{y}}{1 + p_{z_2}} N, \quad F_\theta = \pi_\theta F \quad (1.29)$$

for aggregate investments in securities. Aggregate savings are

$$S = \left[\frac{\delta}{1 + \delta} \frac{\bar{w} - \bar{y}}{1 + p_{z_1}} \left(s_d + \frac{s_f}{p} \right) + \frac{\bar{e}_1}{r} \right] N \quad (1.30)$$

and aggregate current expenditures are

$$E_0 = \left[\frac{1}{1 + \delta} (\bar{w} - \bar{y}) + \bar{e}_0 \right] N. \quad (1.31)$$

1.5 The effect of the skill premium on the sectoral structure

In a general equilibrium, sectoral structure and skill premium are determined simultaneously. As an intermediate step we characterize the sectoral structure as a function of the skill premium and exogenous parameters, keeping in mind that in the end the skill

premium depends on exogenous parameters, too. Not all possible values of skill premia and parameters are of interest, but only those which are reasonable candidates for a general equilibrium, in which both financial sectors are viable, the subsistence of all agents is feasible and a positive skill premium results. The following paragraphs characterize the set of parameter configurations which guarantee these equilibrium properties.

Assumption 1.1 that financial service provision is more skill intensive than goods production ($\kappa_x < k < \kappa_z$) is equivalent to $\frac{\gamma_x}{k} < \omega < \frac{\gamma_z}{k}$ as we know from (1.9). At $\omega_{min} \equiv \frac{\gamma_x}{k}$ the Z -sector vanishes and beyond $\omega_{max} \equiv \frac{\gamma_z}{k}$ there would be no longer an X -sector. Hence, we consider the range $\omega \in (\omega_{min}, \omega_{max})$ in our search for the equilibrium skill premium.

Moreover, according to Fact 1.3, $\omega < \omega_L^+(A_x, A_z, b_L, \bar{e}_0, \frac{\bar{e}_1}{r})$ is required for guaranteeing subsistence for low-skilled agents. $\omega_L^+ \geq \omega_{max}$ holds if A_x, A_z and b_L are large enough (for given $\bar{e}_0, \frac{\bar{e}_1}{r}$), or \bar{e}_0 and $\frac{\bar{e}_1}{r}$ are not too high (for given A_x, A_z, b_L). If $\omega_L^+ < \omega_{max}$, only range $\omega \in (\omega_{min}, \omega_L^+)$ is feasible.

Finally, $\omega \geq b_L/b_H$ is required for $y^H \geq y^L$. This is guaranteed if $\omega_{min} \geq b_L/b_H$, which is equivalent to

$$\gamma_x \geq \frac{\bar{H}}{\bar{L}}.$$

In terms of exogenous fundamentals, the requirements mean that we restrict the possible combinations of exogenous model parameters

$$\xi = \left\{ A_x, A_{z_1}, A_{z_2}, \alpha_x, \alpha_{z_1}, \alpha_{z_2}, b_L, b_H, \bar{H}, \bar{L}, \bar{e}_0, \frac{\bar{e}_1}{r}, \rho, \mu, \delta \right\}$$

to the following set:

$$\Xi_0 \equiv \left\{ \xi \left| \frac{\bar{H}}{\bar{L}} \leq \gamma_x, \frac{\gamma_x}{k} < \tilde{\omega}_{max} \right. \right\}, \quad (1.32)$$

where $k = \frac{b_H \bar{H}}{b_L \bar{L}}$ and $\tilde{\omega}_{max} \equiv \min \left\{ \omega_{max}, \omega_L^+(A_x, A_{z_1}, b_L, \bar{e}_0, \frac{\bar{e}_1}{r}) \right\}$.

In general, the interaction of the allocation of resources between the X -sector and the Z -sector, on the one hand, and the allocation within the Z -sector on Z_1 and Z_2 , on the other hand, are hard to disentangle in an economically transparent way. For qualitatively robust insights into important channels we have to reduce complexity on either the demand or the supply side. In the benchmark analysis presented in Section 1.5.1, 1.5.2 and 1.6, we shut down relative price effects within the financial sector by assuming identical technologies for Z_1 and Z_2 .

Assumption 1.2. $\alpha_{z_1} = \alpha_{z_2} = \alpha_z > \alpha_x$ and $A_{z_1} = A_{z_2} = A_z$.¹²

¹²Without normalization $n_1 = n_2 = 1$, the assumption would read $\frac{A_{z_1}}{n_1} = \frac{A_{z_2}}{n_2}$. That is the provision of financial services per unit of saving must be equal in the two subsectors. For instance, new financial

Assumption 1.2 allows us to put focus on the income effects. In Section 1.8 we consider the case $\alpha_{z_2} > \alpha_{z_1} = \alpha_x$ as a robustness check. Moreover, in the quantitative implementation of the model we solve the model numerically for α_j -values that match U.S. data where $\alpha_{z_2} > \alpha_{z_1} > \alpha_x$.

We analyze first the impact of an increase in the skill premium on structural change within the financial sector.

1.5.1 Within change

The value added in subsector $Z_i, i = \{1, 2\}$, is equal to aggregate expenditure on the produced services. According to (1.28) and (1.29), aggregate expenditures for financial services have the following structure:

$$\frac{p_{z_2}F}{p_{z_1}D} = \frac{s_f \zeta \bar{\eta}}{s_d \bar{\eta} + \frac{1+\delta}{\delta} \frac{\bar{e}_1}{r}} \equiv \Phi(s_d, s_f, \frac{\bar{e}_1}{r}, \zeta(\omega), \bar{\eta}(\omega)) \quad (1.33)$$

where $\zeta(\omega) \equiv \frac{p_{z_2}}{p_{z_1}} \frac{1+p_{z_1}}{1+p_{z_2}}$, $\bar{\eta}(\omega) \equiv \frac{\bar{w}-\bar{y}}{1+p_{z_1}}$ and s_d, s_f are defined in (1.22). While the impacts of saving propensities s_d and s_f on the within structure are straightforward, the role of the skill premium is in general ambiguous. $\zeta(\omega)$ expresses relative price effects. Since $p_{z_1} = p_{z_2} = p_z$ under Assumption 1.2, $\zeta(\omega)$ reduces to one (see discussion in Section 1.8 for the case of changing relative prices within the Z-sector). $\bar{\eta}(\omega)$ is the average supernumerary income weighted by the cost of future subsistence. It captures the income effect on within structural change. If $\bar{e}_1 = 0$, there is no income effect on the demand structure for financial services. For $\bar{e}_1 > 0$, the impact of the skill premium on the value added ratio Φ of sector Z_2 compared to Z_1 depends in the benchmark only on the shape of $\bar{\eta}(\omega)$. The following lemma characterizes the properties of $\bar{\eta}(\omega)$.

Lemma 1.1. *Let exogenous model parameters belong to Ξ_0 defined in (1.32).*

a) *If $\xi \in \Xi_1 \equiv \Xi_0 \cap \{\xi | \alpha_x + \alpha_z > 1\}$, then there exists a threshold $\underline{\omega}(A_x, A_z, k, \frac{b_L \bar{L}}{N}, \bar{e}_0)$ with $\frac{\partial \bar{\eta}}{\partial \omega} \Big|_{\omega=\underline{\omega}} = 0$ so that:*

$$\begin{aligned} \frac{\partial \bar{\eta}}{\partial \omega} &< 0 \text{ for } \omega < \underline{\omega}, \\ \frac{\partial \bar{\eta}}{\partial \omega} &> 0 \text{ for } \omega > \underline{\omega}. \end{aligned}$$

services may be provided more productively than traditional services, but, at the same time, more units of services are needed to transform a unit of saving into future payoff by complex rather than simple financial products.

Especially, define $\Xi_D^1 \equiv \{\xi | \underline{\omega} > \omega_{min}\}$ and $\Xi_D^2 \equiv \{\xi | \underline{\omega} < \tilde{\omega}_{max}\}$. If $\xi \in \Xi_1 - \Xi_D^1$, then $\frac{\partial \bar{\eta}}{\partial \omega} > 0$ for all $\omega \in (\omega_{min}, \tilde{\omega}_{max})$. If $\xi \in \Xi_1 - \Xi_D^2$, then $\frac{\partial \bar{\eta}}{\partial \omega} < 0$ for all $\omega \in (\omega_{min}, \tilde{\omega}_{max})$.

b) For the comparative-static analysis we have:

$$\bar{\eta} \left(\omega \left| A_{x,+}, A_{z,+}, k, \frac{b_L \bar{L}}{N_+}, \bar{e}_0, \frac{\bar{e}_1}{r_-} \right. \right)$$

Proof. Appendix A.3. □

On the one hand, a higher ω raises the average wage. On the other hand, the prices of financial services are increasing, which has a negative effect on the purchasing power. According to Lemma 1.1, the first effect dominates if the skill premium is sufficiently high.

In sum, we have the following partial results about within structural change in the finance sector.

Proposition 1.2. *Let $\bar{e}_1 > 0$.*

- a) A rise in the skill premium leads to structural change from subsector Z_1 to subsector Z_2 (in terms of value added) at high levels of the skill premium ($\omega > \underline{\omega}$) and to structural change from Z_2 to Z_1 at low levels of skill premium.*
- b) For a given skill premium, a rise of $A_x, A_z, k, \frac{b_L \bar{L}}{N}$ or a decline of $\bar{e}_0, \frac{\bar{e}_1}{r}$ lead to structural change from Z_1 to Z_2 . A rise of μ or a decline of ρ also leads to change from Z_1 to Z_2 , even if $\bar{e}_1 = 0$.*

Proof. (1.33), Lemma 1.1 and the fact that a rise in μ or a decline in ρ raises s_f (at cost of s_d). □

The proposition describes only a partial effect. For a full comparative-static equilibrium analysis, we have to combine the direct effects of exogenous fundamentals with their indirect effects through the equilibrium skill premium. We come back on the total effects in Section 1.6.4.

1.5.2 Between change

For $\alpha_{z_1} = \alpha_{z_2} = \alpha_z$ and $A_{z_1} = A_{z_2} = A_z$, aggregate supply of financial services reduces to:

$$Z(= Z_1 + Z_2) = A_z L_z \kappa_z^\alpha.$$

The allocation between the X -and the Z -sector is then determined by the resource constraints:

$$\begin{aligned} a_x^L X + a_z^L Z &= b_L \bar{L}, \\ a_x^H X + a_z^H Z &= b_H \bar{H}. \end{aligned}$$

In an analogous way to (1.14), we get from this as solution:

$$X = \frac{b_L \bar{L}}{a_x^L} \frac{\kappa_z - k}{\kappa_z - \kappa_x}, \quad Z = \frac{b_L \bar{L}}{a_z^L} \frac{k - \kappa_x}{\kappa_z - \kappa_x}. \quad (1.34)$$

Substituting $a_j^L = \frac{1}{A_j \kappa_j^{\alpha_j}}$, we have:

$$\frac{Z}{X} = \frac{A_z}{A_x} \psi(\omega, k), \quad \psi(\omega, k) \equiv \frac{\kappa_z^{\alpha_z}}{\kappa_x^{\alpha_x}} \frac{k - \kappa_x}{\kappa_z - k}. \quad (1.35)$$

This gives us the following result for the comparative-static effects on the supply structure.

13

Proposition 1.3. *An increase in the skill premium shifts the supply structure from goods production to financial services provision. An increase in the high skilled labor share (k) or biased technical change in favor of financial services (so that total factor productivity A_z rises relative to A_x) have the same effect.*

Proof. The signs of the respective partial derivatives in (1.35) follow from $\kappa_z > \kappa_x$ and the Rybczynski analysis. \square

The proposition characterizes the supply structure as a function of exogenous fundamentals and the skill premium. The supply structure interacts with demand, which depends on aggregate income and prices and thus also reacts to the skill premium. To close the analysis, we have to determine the equilibrium skill premium. Section 1.6.3 will summarize the general equilibrium effect of the skill premium on the between sectoral structure.

¹³ Note that (1.35) characterizes the supply structure of labor produced output. If capital is used as set-up capital as in the extended model in Section 1.7.5, then X is indeed the total size of final output in the goods sector. In the baseline model considered here there is in addition the output generated for old age consumption by past capital investments. Thus, the total size of goods transactions becomes $\bar{X} \equiv X + rD + \mu RF$ with $X = E_0 + S = E_0 + D + F$ and the between structural change ratio is $\bar{\Psi} \equiv \frac{p_z D + p_z F}{\bar{X}} = \frac{p_z D + p_z F}{X + rD + \mu RF}$ with D, F, E_0 and S from (1.28)-(1.31). It is, ceteris paribus, increasing in ω if $S'E_0 - SE'_0 - (\mu R - r)(DF' - FD') > 0$ where D', F', S' and E'_0 are the respective derivatives with respect to ω . This means, if the between change ($S'E_0 - SE'_0$) is larger than within change ($DF' - FD'$) multiplied with the return difference ($\mu R - r$).

1.6 General equilibrium

Aggregate demand in the X -sector is composed of consumer goods demand, E_0 , and investment goods demand, $S = D + F$. On top of it, old agents consume the output generated by the capital they invested in the period before.

Aggregating the individual budget constraints (1.2), we obtain:

$$E_0 + D + F + p_{z_1}D + p_{z_2}F = W, \quad (1.36)$$

where $W = W_x + W_z$, $W_x = X$ and $W_z = p_{z_1}G^{z_1}(H_{z_1}, L_{z_1}) + p_{z_2}G^{z_2}(H_{z_2}, L_{z_2})$. If the Z_1 and Z_2 -markets are cleared, we have $G^{z_1}(H_{z_1}, L_{z_1}) = D$ and $G^{z_2}(H_{z_2}, L_{z_2}) = F$ so that (1.36) reduces to

$$E_0 + D + F = X.$$

Thus, the goods market is automatically cleared if the markets for financial services are cleared.

Aggregate demand for financial services comes from savings in deposits D and savings in securities F . Adding up (1.28) and (1.29), we have for aggregate demand in the Z -sector

$$Z^D = \left(\frac{\delta}{1 + \delta} \frac{\bar{w} - \bar{y}}{1 + p_z} + \frac{\bar{e}_1}{r} \right) N. \quad (1.37)$$

From (1.34) we know that aggregate Z -supply in production equilibrium is

$$Z^S = A_z b_L \bar{L} \kappa_z^{\alpha_z} \frac{k - \kappa_x}{\kappa_z - \kappa_x} \quad (1.38)$$

where $a_z^L = \frac{1}{A_z \kappa_z^{\alpha_z}}$ was used.

1.6.1 Existence, uniqueness and stability of equilibrium

Both market sides are functions of ω (which works through \bar{w} and p_z on the demand side and through skill intensities κ_x, κ_z on the supply side). For a stable equilibrium, the condition

$$\frac{dZ^D}{d\omega} < \frac{dZ^S}{d\omega} \quad (1.39)$$

is required at the market clearing ω -value. Since p_z is increasing in ω , inequality (1.39) guarantees that a rise in price p_z goes hand in hand with a reduction of excess demand and a fall in the price reduces excess supply.

The supply function is characterized by the following fact.

Fact 1.6. Z^S is an increasing strictly concave function of ω starting at $\lim_{\omega \rightarrow \omega_{min}} Z^S = 0$ and approaching $A_z b_L \bar{L} k^{\alpha_z}$ at ω_{max} . More specifically,

$$Z^S = A_z b_L \bar{L} \frac{\gamma_z^{\alpha_z}}{\gamma_z - \gamma_x} g(\omega, k), \quad g(\omega, k) = \omega^{-\alpha_z} (k\omega - \gamma_x). \quad (1.40)$$

Proof. Appendix A.3. □

For the demand side the following fact applies.

Fact 1.7. Aggregate demand for financial services is given by:

$$Z^D = \left[\frac{\delta}{1 + \delta} \bar{\eta} \left(\omega \left| A_x, A_z, k, \frac{b_L \bar{L}}{N}, \bar{e}_0, \frac{\bar{e}_1}{r} \right|_{\substack{+ \\ + \\ + \\ + \\ - \\ -}} \right) + \frac{\bar{e}_1}{r} \right] N,$$

where $\bar{\eta}$ was discussed in Lemma 1.1. For all $\xi \in \Xi_1$, Z^D is defined and positive on the ω -domain $(\omega_{min}, \tilde{\omega}_{max})$. Moreover, it is either U-shaped in ω (for $\xi \in \Xi_D \equiv \Xi_1 \cap \Xi_D^1 \cap \Xi_D^2$), increasing over the entire domain (for $\xi \in \Xi_1 - \Xi_D^1$) or declining for all ω (if $\xi \in \Xi_1 - \Xi_D^2$).

Proof. Equation (1.37) and Lemma 1.1. □

Figure 1.1 shows in the (ω, Z) -space the supply and demand curves under the assumption that

$$Z^D(\tilde{\omega}_{max}) < Z^S(\tilde{\omega}_{max}), \quad (1.41)$$

where $\tilde{\omega}_{max}$ was defined in (1.32).¹⁴

If inequality (1.41) holds, then the market clearing condition $Z^D(\omega) = Z^S(\omega)$ has a unique solution ω^* within $(\omega_{min}, \tilde{\omega}_{max})$. Moreover, stability condition (1.39) is fulfilled at ω^* . This establishes the following proposition.

Proposition 1.4. Define $\Xi_E = \Xi_1 \cap \{\xi | Z^D(\tilde{\omega}_{max}) < Z^S(\tilde{\omega}_{max})\}$. For all $\xi \in \Xi_E$, there exists a unique and stable equilibrium.

Proof. Continuity of Z^D on $\omega \in (\omega_{min}, \tilde{\omega}_{max})$ and properties of the shape of Z^D established in Fact 1.7. □

1.6.2 Equilibrium skill premium

For the comparative-static equilibrium analysis, we have to look at the excess demand function $Z^D - Z^S$. Because of stability condition $\frac{\partial(Z^D - Z^S)}{\partial \omega} < 0$, we know that for any

¹⁴If $\tilde{\omega}_{max} = \omega_L^+$, then $Z^D(\tilde{\omega})$ is to be read as $Z^D(\omega) < Z^S(\omega)$ for all $\omega < \omega_L^+ - \epsilon$, with ϵ arbitrarily small. Figure 1.1 assumes $\xi \in \Xi_D$; yet, from Fact 1.7 it is obvious that for $\xi \in \Xi_1 - \Xi_D^1$ the Z^D -curve would cross the Z^S -curve at ω^* as in Case I, whereas for $\xi \in \Xi_1 - \Xi_D^2$ we would have at ω^* the situation illustrated in Case II.

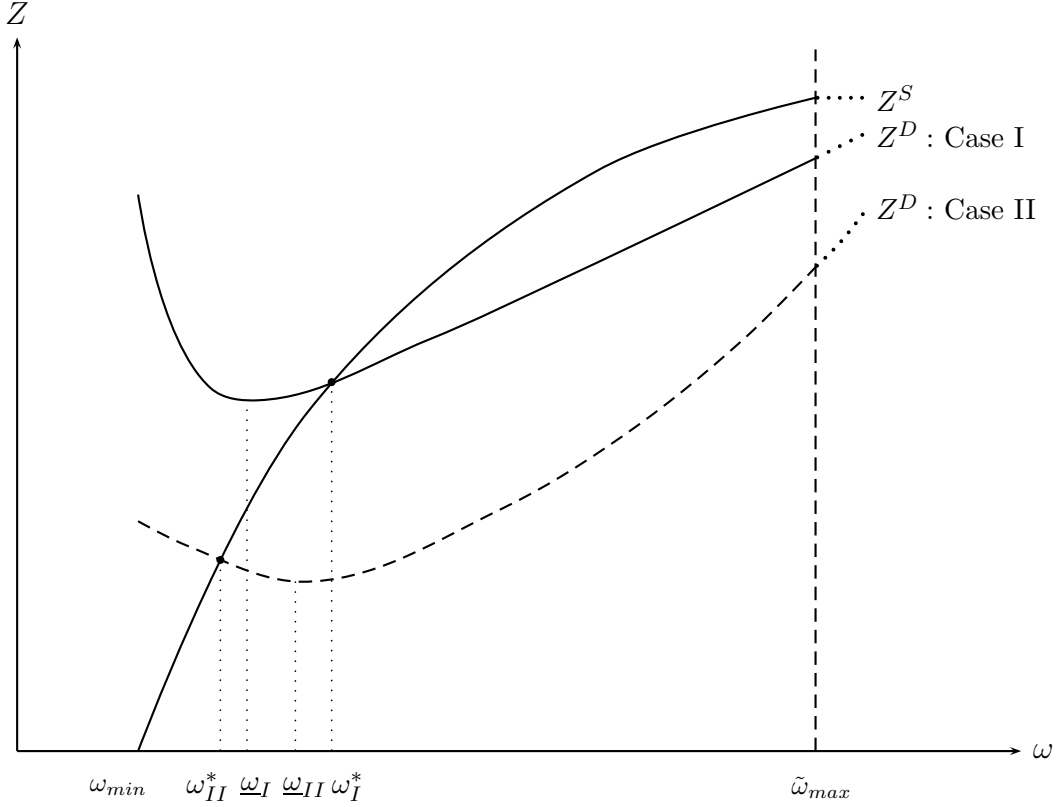


Figure 1.1: Equilibrium in the financial service sector

exogenous change of a component i of $\xi \in \Xi_E$

$$\text{sign} \frac{\partial \omega^*}{\partial i} = \text{sign} \frac{\partial (Z^D - Z^S)}{\partial i} \Big|_{Z^D=Z^S}.$$

For signing the impact of exogenous fundamentals on the equilibrium, we express excessive demand explicitly as a function of model parameters. Using (1.26) and (1.12), we have

$$\frac{\bar{w}N}{1+p_z} = A_x b_L \bar{L} D_1(\omega | \frac{A_z}{A_x} \underset{+}{+}, k), \quad (1.42)$$

where $D_1 \equiv \frac{\Gamma_x(1+\omega k)}{\omega^{\alpha_x} + \frac{A_x \Gamma_x}{A_z \Gamma_z} \omega^{\alpha_z}}$ and the signs below parameters in (1.42) express the signs of their impact on D_1 . Term D_1 captures the purchasing power effect.

Moreover, substituting (1.12) for p_{z_1} in (1.18) we can write the term $\frac{\delta}{1+\delta} \frac{\bar{y}}{1+p_z} - \frac{\bar{e}_1}{r}$ in the form:

$$D_0(\omega | \frac{A_z}{A_x} \underset{+}{+}, \bar{e}_0, \frac{\bar{e}_1}{r}) = \frac{1}{1+\delta} \left[\frac{\delta \bar{e}_0}{1 + \frac{A_x \Gamma_x}{A_z \Gamma_z} \omega^{\alpha_z - \alpha_x}} - \frac{\bar{e}_1}{r} \right]. \quad (1.43)$$

Term D_0 captures the effect of the subsistence requirements on the aggregate demand

for financial services. The sign of the square-bracketed term is positive if the present subsistence expenditure \bar{e}_0 dominates the future subsistence expenditure \bar{e}_1 . It is negative if \bar{e}_1 dominates \bar{e}_0 . For the economic interpretation of the relevant notion of dominance it is useful to recall $\frac{A_x \Gamma_x}{A_z \Gamma_z} \omega^{\alpha_z - \alpha_x} = p_z$. Thus, $D_0(\omega | \frac{A_z}{A_x}, \bar{e}_0, \frac{\bar{e}_1}{r}) > 0$ ($=, < 0$) if and only if

$$\frac{\delta \bar{e}_0}{1 + p_z} > \frac{\bar{e}_1}{r} \quad (=, < \frac{\bar{e}_1}{r}, \text{ resp.}). \quad (1.44)$$

This is exactly the condition for a rising (constant, declining, resp.) saving rate derived in Fact 1.4.b). (Note that $p = 1$ in the benchmark case.) If present subsistence expenditures are more pressing than future ones, people save more and demand more financial services if they become richer and get farther away from subsistence problems.

Using D_0 and (1.42) in (1.37) and combining the result with (1.40), we conclude that $Z^D - Z^S$ is equal to the term

$$A_x b_L \bar{L} \left[\frac{\delta}{1 + \delta} D_1(\omega | \frac{A_z}{A_x}, \frac{k}{+}) - \frac{N}{A_x b_L \bar{L}} D_0(\omega | \frac{A_z}{A_x}, \bar{e}_0, \frac{\bar{e}_1}{r}) - \frac{A_z}{A_x} \frac{\gamma_z^{\alpha_z}}{\gamma_z - \gamma_x} g(\omega, k) \right]. \quad (1.45)$$

Hence, \bar{e}_1 has a positive impact on $Z^D - Z^S$ and thus on ω^* ; \bar{e}_0 has a negative impact. $\frac{A_z}{A_x}$ and k have opposing effects so that their impacts cannot be signed unambiguously by inspection of (1.45).

The most interesting question is how technical change affects the equilibrium skill premium. For this we have to look at the impact of $\frac{A_x b_L \bar{L}}{N}$ on $Z^D - Z^S$. (Since $\frac{A_z}{A_x}$ has an ambiguous effect, we only consider uniform progress across sectors, that is, total factor productivity A_z rises *pari passu* with A_x .) The answer depends on condition (1.44). If $\frac{\delta \bar{e}_0}{1 + p_z} > \frac{\bar{e}_1}{r}$, D_0 is positive and ω^* increases if $\frac{A_x b_L \bar{L}}{N}$ rises. If $\frac{\delta \bar{e}_0}{1 + p_z} < \frac{\bar{e}_1}{r}$, then D_0 is negative and ω^* declines if $\frac{A_x b_L \bar{L}}{N}$ increases. For $\bar{e}_0 = \bar{e}_1 = 0$, $\frac{A_x b_L \bar{L}}{N}$ has no effect.

In sum, we have the following partial effects of the parameters on the equilibrium skill premium:¹⁵

$$\omega^* \left(\frac{A_z}{A_x}, k, \frac{A_x b_L \bar{L}}{N}, \bar{e}_0, \frac{\bar{e}_1}{r} \right), \quad (1.46)$$

$\begin{matrix} ? \\ +/- \end{matrix}$

where the impact of $\frac{A_x b_L \bar{L}}{N}$ depends on the cases discussed above.

All addressed effects refer to the partial derivatives, that is, they hold under the condition that other parameters do not change simultaneously. Economically this means, the effects come from a single source. In particular, for the effect of $\frac{b_L \bar{L}}{N}$ on ω^* , skill richness

¹⁵The signs below the parameters represent the partial derivatives. The combination $+/-$ is used for pointing to case-dependent impacts. A question mark means that the impact of the respective parameter cannot be signed without further investigation.

$k = \frac{b_H \bar{H}}{b_L \bar{L}}$ is held constant in the comparison. This requires a careful interpretation of the described effect of $\frac{b_L \bar{L}}{N}$. The following fact provides an economically meaningful description of the variations which are consistent with a constant k and a rise in $\frac{b_L \bar{L}}{N}$.

Fact 1.8. *A rise in $\frac{b_L \bar{L}}{N}$ is consistent with a constant k if there is:*

- a) *Uniform factor-augmenting technical progress, raising b_L pari passu with b_H .*
- b) *A shift in labor supply from unskilled to skilled labor accompanied by factor augmenting progress that is biased towards the low-skilled. (Note that such low-skilled labor augmentation depresses the relative wage of the unskilled – like skill-biased technical change.)*

Proof. Use $N = \bar{L} + \bar{H}$ for $\frac{N}{b_L \bar{L}} = \frac{1 + \frac{\bar{H}}{\bar{L}}}{b_L}$. Hence, $k = \frac{b_H \bar{H}}{b_L \bar{L}}$ remains constant under a decrease in $\frac{N}{b_L \bar{L}}$ if either b_L and b_H rise proportionally and \bar{H}/\bar{L} does not change or $\frac{\bar{H}}{\bar{L}}$ rises and b_L rises such that $\frac{b_L}{b_H}$ grows proportionally to $\frac{\bar{H}}{\bar{L}}$. \square

With these clarification the following proposition summarizes the comparative-static equilibrium results.

Proposition 1.5. *Let $\bar{e}_0 > 0$ or $\bar{e}_1 > 0$.*

- a) *Uniform productivity growth across sectors (raising A_x and A_z proportionally) or uniform factor-augmenting technical progress (raising b_L and b_H proportionally) have a positive effect on the equilibrium skill premium if the present subsistence expenditure dominates the future subsistence expenditure; if the future subsistence expenditure dominates, then the skill premium declines.*
- b) *A shift of labor supply from unskilled to skilled work accompanied by factor augmentation which is biased towards low-skilled labor has the same effect on the equilibrium skill premium as factor augmenting progress that is uniform.*
- c) *The equilibrium skill premium rises, if future subsistence expenditure (\bar{e}_1) increases or present subsistence expenditure (\bar{e}_0) declines.*

Proof. Fact 1.8 and main text. \square

For the economic intuition behind a) and b) it is useful to remember Fact 1.4.b). If present subsistence expenditure weighs more than future subsistence requirements then the saving rate and therefore demand for financial services are rising with income. Since the financial services are more skill intensive than goods, this rise of demand induces a rise in the skill premium. The rising income in turn comes from technical progress or a better educated workforce. The intuition for c) is: If future subsistence expenditure is

high, agents have to save more and need more financial services; and if present subsistence expenditure is low, they can afford to save more and to spend more for financial services.

It is worth noting that positive subsistence expenditure ($\bar{e}_0 > 0$ or $\bar{e}_1 > 0$) is essential for the comparative-static results stated in Proposition 1.5. For $\bar{e}_0 = \bar{e}_1 = 0$, expression (1.45) boils down to

$$A_x b_L \bar{L} \left[\frac{\delta}{1 + \delta} D_1 \left(\omega \left| \frac{A_z}{A_x}, k_+ \right. \right) - \frac{A_z}{A_x} \frac{\gamma_z^{\alpha_z}}{\gamma_z - \gamma_x} g \left(\omega, k_+ \right) \right].$$

Thus, uniform productivity growth has no effect in this case nor has $\frac{b_L \bar{L}}{N}$.

1.6.3 Structural change between production and financial service sectors

Combining the results of subsections 1.6.2 and 1.5.2, we obtain the following results for the structural change between production and financial services in equilibrium:

Proposition 1.6. *For all $\xi \in \Xi_E$, at given $\frac{A_z}{A_x}$, k , any change in other exogenous fundamental which raises (lowers) the skill premium leads to structural change from X to Z (Z to X , respectively).*

Proof. Equation (1.35). Since p_z rises with ω , the rise of ψ immediately implies that $\frac{p_z Z}{X}$ rises too. \square

1.6.4 Structural change within the financial sector

Finally, for structural change within the financial sector, we have the following results in equilibrium:

Proposition 1.7. *Let $\underline{\omega}$ be the threshold defined in Lemma 1.1 and parameters fulfill $\xi \in \Xi_E$. Then, under the assumption that prices do not differ across financial services, the following comparative-static results hold for structural change within the financial sector as long as $\bar{e}_1 > 0$:*

- a) *At high levels of the skill premium ($\omega^* > \underline{\omega}$), a fall of \bar{e}_0 leads to a shift from Z_1 to Z_2 . In addition, if present subsistence expenditure dominates future subsistence expenditure, uniform productivity growth across sectors (i.e., a proportional rise of A_x and A_z) as well as an increase in $\frac{b_L \bar{L}}{N}$ change the structure within the financial sector from Z_1 towards Z_2 . According to Proposition 1.5 and 1.6, these changes induce an increase in the inequality level ω^* , accompanied by a simultaneous structural change from the goods to the financial service sector.*

- b) At low levels of the skill premium ($\omega^* < \underline{\omega}$), a fall of \bar{e}_1 leads to a shift from Z_1 to Z_2 . In addition, if future subsistence expenditure dominates present subsistence expenditure, uniform productivity growth across sectors as well as and an increase in $\frac{b_L \bar{L}}{N}$ change the structure within the financial sector from Z_1 towards Z_2 . However, according to Proposition 1.5 and 1.6, these changes correspond to a decrease in the inequality level ω^* , accompanied by a simultaneous a structural change from the financial service to the goods sector.
- c) Financial product innovation (a rise of μ) or rising attractiveness of risky investments (a decline of ρ) lead to structural change from Z_1 to Z_2 , even if $\bar{e}_1 = 0$.

Proof. Using (1.33), (1.46), and Lemma 1.1, we have

$$\Phi \left\{ s_d, s_f, \frac{\bar{e}_1}{r}, \bar{\eta} \left[\omega^* \left(\frac{A_z}{A_x}, k, \frac{A_x b_L \bar{L}}{N}, \bar{e}_0, \frac{\bar{e}_1}{r} \right), A_x, A_z, k, \frac{b_L \bar{L}}{N}, \bar{e}_0, \frac{\bar{e}_1}{r} \right] \right\},$$

where the signs below the parameters show the sign of the respective partial derivative of the functions $\Phi\{\cdot\}$, $\bar{\eta}[\cdot]$ and $\omega^*(\cdot)$. The plus below ω^* applies for $\omega^* > \underline{\omega}$, the minus for $\omega^* < \underline{\omega}$. The plus below $\frac{A_x b_L \bar{L}}{N}$ applies for the case that \bar{e}_0 dominates \bar{e}_1 ; the minus applies if \bar{e}_1 dominates \bar{e}_0 . For the impacts of μ and ρ note that s_f is rising and s_d is declining in μ and rising in ρ . \square

It is worth noting that for $\bar{e}_1 = 0$ there is no income effect on the portfolio structure so that the channel between skill premium and financial structure is shut down. Since in the benchmark considered here relative price effects within the financial sector were shut down too, for $\bar{e}_1 = 0$ only financial innovation (a rise in μ) and rising relative returns on risky investment (a decline of ρ) remain as sources of structural change within the financial sector. This will change in the model variant with different technologies for Z_1 and Z_2 considered in Section 1.8.

The punchline of the general equilibrium analysis in the baseline model is: When the skill premium has reached a certain level, a rise in average income leads to rising inequality and to two fold structural change towards and within the financial sector simultaneously. The rise in income can be triggered by a general rise of productivity or by an increased selection of the population into higher education (accompanied by labor augmenting progress that makes low-skilled labor abundant relative to skilled labor). The income effects generated by technical progress or education are robust drivers of the developments outlined at the beginning of this paper. They can explain a rising skill premium and the twofold structural change towards and within finance by a single source, holding everything else constant. Yet, of course, in reality the effects triggered by this source are

overlaid by many other things that happen at the same time. The model points to a series of other exogenous fundamentals that affect skill premium and economic structure. Thus, the specific combination of determinants that actually determine the observed patterns of inequality and structural change can only be identified by empirical analysis. The quantitative analysis in Section 1.9 illustrates possible combinations of exogenous factors which are consistent with the development observed from 1980 onwards. Before turning to the quantitative analysis, we want to add realism to the model by considering extensions which account for important aspects that have been neglected in the “perfect-world” analysis presented so far.

1.7 Extensions

Five extensions are considered: Fixed costs in the financial sector, rents in the financial sector, distorted portfolio choices of households, participation constraints in finance sub-sector Z_2 and set-up capital for firms. Like the equilibrium analysis in the benchmark, the extended analysis is based on Assumption 1.2. Moreover, for avoiding too many case distinctions, dominance of \bar{e}_0 over \bar{e}_1 is assumed in this section.

1.7.1 Fixed costs in the financial sector

Suppose that financial services are provided by banks. A bank b , serving N_b clients, needs $K_b = f_B N_b$ units of goods to set up the capacity to serve them. We assume that the fixed cost K_b is financed by a lump-sum fee

$$\tau = f_B$$

imposed on the clients. That is, bank size and number of banks affect neither aggregate fixed costs

$$K_B = f_B N$$

nor the households’ budget constraint. In the latter, y^l reduces to $y^l - \tau$ so that the supernumerary budget becomes $y^l - \bar{y}_+$, with $\bar{y}_+ = \bar{y} + \tau = \bar{e}_0 + f_B + (1 + p_{z1})\bar{e}_1/r$.

Hence, fixed cost f_B has the same comparative-static effect on household choices as an increase in subsistence expenditure \bar{e}_0 . For the X -market this means, on the one hand, the absorption of X by households’ consumption and investment is reduced by $K_B = f_B N$. On the other hand, K_B is spent by banks to set up the capacity to serve their clients. In sum, we have

$$E_0 - f_B N + D + F + K_B = X$$

for the goods market clearing, which reduces to the condition in the benchmark model:

$$E_0 + D + F = X$$

since $f_B N = K_B$. Hence, goods markets are cleared whenever the Z -markets are cleared.

In the markets for financial services, demand is reduced by the fact that $\bar{w} - \bar{y}_+$ rather than $\bar{w} - \bar{y}$ is now the relevant supernumerary income. The supply side remains unaffected. In equilibrium, the implications of fixed costs can be derived by looking in the benchmark model at the effect of a rise of \bar{e}_0 to $\bar{e}_0 + f_B$.

Proposition 1.8. *A decline in fixed costs f_B has the following effects:*

- a) *The skill premium rises.*
- b) *The between sectoral structure shifts from X to Z .*
- c) *The within sectoral structure shifts from Z_1 to Z_2 at high levels of the skill premium ($\omega^* > \underline{\omega}$). At low levels of the skill premium ($\omega^* < \underline{\omega}$) the effect is ambiguous.*

Proof. Comparative-static results for \bar{e}_0 in Proposition 1.5-1.7. □

1.7.2 Rents in the financial sector

Suppose that a club of agents in the finance sector has the power to extract rents from financial service provision.¹⁶ One may think of rentiers who have unearned property rights or an elite subgroup of employees in the financial sector. We make two crucial assumptions: First, whoever are the rent extracting agents, they spend the rent like other agents. Thus, the redistribution of rents has no income effect on aggregate demand. (Total subsistence requirements and aggregate supernumerary income remain unchanged). Second, nobody can enter the club from outside so that the rent does not affect labor allocation.

In the presented model, two instruments can be used to extract rents. First, a fixed fee $\tilde{\tau}$ as in extension 1.7.1, but with:

$$\tilde{\tau} > f_B.$$

Aggregate rents $(\tilde{\tau} - f_B)N$ are lump-sum redistributed. Everybody pays $\tilde{\tau}$ and an elite N_0 receives the rent. Thus, average supernumerary income becomes

$$\bar{w} - \bar{y} - \tilde{\tau} + \frac{N_0}{N} \frac{(\tilde{\tau} - f_B)N}{N_0} = \bar{w} - \bar{y} - f_B.$$

¹⁶As pointed out in the introduction, there is robust evidence that indeed a substantial finance premium exists. This paper deals with the consequences of rents, not with possible explanations why they exist.

In this case, the rent has no effects on aggregate income, expenditure structure, labor allocation, relative prices or the skill premium. Nevertheless, there is lump-sum redistribution of income from the real to the financial sector and within the financial sector. This redistribution implies for the sectoral income shares:

$$\frac{p_z Z + (\tilde{\tau} - f_B) N}{X}$$

and

$$\frac{p_z F + \nu(\tilde{\tau} - f_B) N}{p_z D + (1 - \nu)(\tilde{\tau} - f_B) N},$$

respectively, where ν is the share of the elite rent going to new finance. It is obvious that a rising finance rent increases the total finance share in the economy. For a given rent distribution ν , a rise in $\tilde{\tau}$ raises the income share of new finance relative to traditional finance as long as $\nu D > (1 - \nu)F$, that is as long as the new finance share is not too large. A rise in ν trivially leads to a rise in the new finance share.

A second instrument of rent extraction would be to charge a markup on unit cost prices in the financial sector so that households have to pay

$$\tilde{p}_{z_i} = p_{z_i}(1 + o_i)$$

for financial services.

Using (1.12), we have

$$\tilde{p}_{z_i} = (1 + o_i) \frac{A_x}{A_{z_i}} \frac{\Gamma_x}{\Gamma_{z_i}} \omega^{\alpha_{z_i} - \alpha_x}.$$

In the benchmark case with $p_{z_1} = p_{z_2}$ a rent $o_1 = o_2 = o$ decreases D_1 in (1.42) to $\frac{A_x \Gamma_x (1 + \omega k)}{\omega^{\alpha_x} + \frac{(1+o)A_x \Gamma_x}{A_z \Gamma_z} \omega^{\alpha_z}}$ and decreases D_0 in (1.43) to $\frac{1}{1+\delta} \left[\frac{\delta \bar{e}_0}{1 + (1+o) \frac{A_x \Gamma_x}{A_z \Gamma_z} \omega^{\alpha_z - \alpha_x}} - \frac{\bar{e}_1}{r} \right]$. Hence, o has an ambiguous impact on $Z_D - Z_S$ and thus on ω^* .

Proposition 1.9. *Rents in the financial sector have the following effects:*

- a) *If rents are extracted by lump sum fees, they have no allocative equilibrium effects. Yet, there is a redistributive effect that raises the finance share in total income. The structure of the subsector shares within finance depends on how the earned rents are distributed on traditional and new finance, respectively.*
- b) *If rents are extracted by a markup on financial service prices, there is a redistributive effect towards (and within) the financial sector. Yet, the markups affect all equilibrium values in a generally ambiguous way.*

Proof. Main text. □

1.7.3 Distorted portfolio choice

Several empirical studies have pointed out that people get confused in dealing with complex financial markets (see C  lerier and Vall  e (2014) and the literature discussed there). In our model, the complex part that households have to solve is the choice of the portfolio of the securities. The choice may be based on a wrong assessment of relative risks and returns of different securities. In this case, we have distortion within Z_2 and consumption levels planned for the future may be deceived by actual payoffs.¹⁷ As our study focuses on structural change between X and Z as well as between Z_1 and Z_2 , we do not consider such distortions here. Rather we focus on distortions coming from misperception of the opportunities to save by securities investment rather than in deposits.

In particular, people may have wrong beliefs $\tilde{\mu}$ about the measure of future environments covered by state-contingent securities, relative to the non-covered part of possible future events. They may also misjudge the relative payoff of deposits compared to the payoffs of securities and base their decisions on a distorted $\tilde{\rho}$. Such distortions affect the propensities to save in deposits and in securities. For instance, if agents are euphoric about investments in securities and believe that $\tilde{\mu} > \mu$ or $\tilde{\rho} < \rho$, then s_f rises while s_d declines. The total propensity to save, however, does not change in the benchmark model with $p_{z_1} = p_{z_2}$.¹⁸ Therefore, the only consequence of $\tilde{\mu} > \mu$ or $\tilde{\rho} < \rho$ is sectoral change within the financial sector. According to (1.33), Φ rises.

Proposition 1.10. *Euphoric beliefs about measure or performance of state-contingent financial instruments lead to within sectoral change from Z_1 to Z_2 . Equilibrium skill premium and (X, Z) -structure are not affected in the benchmark model (with identical technologies in Z_1 and Z_2).*

Proof. Equation (1.33). □

1.7.4 Participation constraints

Suppose that a fixed fee τ is charged only to agents who invest in securities. Moreover, assume that there is a participation constraint:

$$\begin{aligned} y^L &> \bar{y} > y^L - \tau, \\ y^H &> y^H - \tau > \bar{y}. \end{aligned}$$

¹⁷Falkinger (2014) focuses on such distortions in an one sector economy.

¹⁸For $p_{z_1} \neq p_{z_2}$, however, we would have $s_d + \frac{s_f}{p}$ for the marginal propensity to save, as shown by (1.23). Thus, μ and ρ impact also on Z^D and therefore on ω and all other equilibrium outcomes. See Section 1.8 for a more detailed discussion.

Then, low-skilled agents do not participate in the securities market, while high-skilled agents do. According to (A.18) in Appendix A.2, we have for $l = L$:

$$s^L = d^L = \frac{\delta}{1 + \delta} \frac{y^L - \bar{y}}{1 + p_z} + \frac{\bar{e}_1}{r}.$$

For $l = H$, saving behavior is given by (1.20) and (1.21) with $\bar{y}_+ = \bar{y} + \tau$.

This gives us the following aggregate saving levels:

$$\begin{aligned} D &= \frac{\delta}{1 + \delta} \frac{1}{1 + p_z} \left[(y^L - \bar{y}) \bar{L} + s_d (y^H - \bar{y}_+) \bar{H} \right] + \frac{\bar{e}_1}{r} N \\ F &= s_f \frac{\delta}{1 + \delta} \frac{\bar{H}}{1 + p_z} (y^H - \bar{y}_+) \\ S &= \left(\frac{\delta}{1 + \delta} \frac{\bar{w} - \bar{y} - \tau \frac{\bar{H}}{N}}{1 + p_z} + \frac{\bar{e}_1}{r} \right) N. \end{aligned}$$

Comparing S with Z^D in (1.37), we see that fee τ , combined with the participation constraint, impacts on Z^D and thus on the skill premium and the (X, Z) -structure like an increase of \bar{e}_0 to

$$\tilde{e}_0 = \bar{e}_0 + \tau \frac{\bar{H}}{N}.$$

Moreover, $\frac{F}{D} = \frac{s_f \bar{H}}{\frac{(y^L - \bar{y}) \bar{L}}{y^H - \bar{y}_+} + s_d \bar{H} + \frac{1 + \delta}{\delta} \frac{(1 + p_z) \bar{e}_1}{r} \frac{N}{y^H - \bar{y}_+}}$ is declining in τ . Thus, the participation constraint does not change the comparative-static effects of fixed cost τ described in Proposition 1.8.

The above conclusion is only valid if τF is absorbed by real fixed cost requirements as discussed in Section 1.7.1. If τF is a rent which is redistributed back to high-skilled agents, as in Section 1.7.2, we have $(y^H - \bar{y} - \tau) \bar{H} + \tau \bar{H} = y^H - \bar{y}$ instead of $y^H - \bar{y}_+$ so that

$$\begin{aligned} D &= \frac{\delta}{1 + \delta} \frac{\bar{w} - \bar{y}}{1 + p_z} N (1 - s_f \beta_H) + \frac{\bar{e}_1}{r} N \\ F &= s_f \frac{\delta}{1 + \delta} \frac{\bar{H}}{1 + p_z} (y^H - \bar{y}) \\ S &= \left(\frac{\delta}{1 + \delta} \frac{\bar{w} - \bar{y}}{1 + p_z} + \frac{\bar{e}_1}{r} \right) N \end{aligned}$$

with $\beta_H \equiv \frac{y^H - \bar{y}}{\bar{w} - \bar{y}} \frac{\bar{H}}{N}$ denoting the income share of high-skilled agents. For the high-skilled nothing changes, but the low-skilled are only saving through D . This means that, compared to the benchmark, we have an increase in D and a decrease in F . $Z^D = S$ coincides with the expression in (1.30) so that equilibrium skill premium and (X, Z) -structure are

not changed compared to the baseline.¹⁹

For the within sectoral structure in the Z -sector, we have in the benchmark case with $p_{z_1} = p_{z_2} = p_z$:

$$\begin{aligned}\frac{F}{D} &= \frac{s_f \beta_H}{1 - s_f \beta_H + \frac{1+\delta}{\delta} \frac{1+p_z}{\bar{w}-\bar{y}} \frac{\bar{e}_1}{r}} \\ &= \frac{s_f \beta_H \bar{\eta}}{s_d \bar{\eta} + s_f (1 - \beta_H) \bar{\eta} + \frac{1+\delta}{\delta} \frac{\bar{e}_1}{r}} \equiv \tilde{\Phi}\end{aligned}$$

Comparing this with (1.33), we conclude that $\tilde{\Phi} < \Phi$ because $s_f(1 - \beta_H) > 0$. Yet, the proportion of total expenditure on new finance relative to expenditure on traditional finance $\frac{p_z F + \tau \bar{H}}{p_z D} = \frac{F}{D} + \frac{\tau \bar{H}}{p_z D}$ is ambiguous. Rent τ increases the new finance share, but the participation constraint induces a shift of the portfolio towards safe assets.

1.7.5 Set-up capital for firms

In the baseline model invested capital is transformed by linear technologies, using capital as the only input, into future output. The extension in this section shows that the baseline can be seen as kind of reduced form of a richer model, in which capital is needed to set up firms. We assume now that firms in the X -sector use capital to set up the technology G^x which then produces output by employing low-skilled and high-skilled labor. Each established firm $\nu \in \{1, \dots, M\}$ produces a variety $x_\nu = G^x(L_{x_\nu}, H_{x_\nu})$ under monopolistic competition with free entry. Consumers spend the supernumerary income $e_t - \bar{e}_t$ according to a CES-utility function with substitution elasticity $\sigma > 1$ symmetrically over the variants x_ν in the X -sector, which implies an instantaneous indirect utility function of the form $\log(e_t - \bar{e}_t)$ (see Section 1.7.5.1) like before. So saving decision and portfolio choice remain the same as in the baseline model. Firms have positive operating profits which are distributed as payoff to the investors (see Section 1.7.5.2).

1.7.5.1 Consumer problem

Let the instantaneous utility of households be given by $u = \left[\sum_{\nu=1}^M x_\nu^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$, $\sigma > 1$. Then, prices are determined by a constant markup on unit cost of production

$$p_\nu = \frac{\sigma}{\sigma - 1} c(w_H, w_L), \quad (1.47)$$

¹⁹For $p_{z_1} \neq p_{z_2}$, however, the change in Z_2^D would also affect ω and all other equilibrium outcomes.

where $c(w_H, w_L)$ are the unit costs (as in Section 1.2) and w_H, w_L are the factor prices. Moreover, demand for variety x_ν of a household that spends “supernumerary budget” $e - \bar{e}$ is

$$x_\nu = (e - \bar{e}) \frac{p_\nu^{-\sigma}}{P^{1-\sigma}}, \quad P \equiv \left[\sum_{\nu=1}^M p_\nu^{1-\sigma} \right]^{\frac{1}{1-\sigma}}.$$

Since product variants use identical production technologies, their unit cost and prices are identical, too. Thus, x_ν reduces to $x = \frac{e - \bar{e}}{p_\nu M}$. Using this in u , we obtain for the instantaneous indirect utility $u = \frac{e - \bar{e}}{P}$. We set the price as numéraire (i.e., $p_\nu = 1$) so that the variety effect is $P = M^{\frac{1}{1-\sigma}}$. Due to the log-specification, this variety effect, though affecting the level of utility, does not matter for the intertemporal decision.²⁰ Thus, $\max \mathbb{E} \log(u) = \max \mathbb{E} \log(e_t - \bar{e}_t)$, which is identical to the intertemporal problem in Section 1.2.2.

1.7.5.2 Firm entry and production in the X -sector

There are two types of set-up technologies, which induce capital demand of firms: A robust set-up technology which requires c_0 units of capital. Firms set up by the robust technology will be producing tomorrow under any condition (i.e., in Θ and $\bar{\Theta}$). Furthermore, there are risky set-up technologies with set-up input c_θ , which are only effective if state $\theta \in \Theta$ occurs. Otherwise, their set-up fails. In an analogous way to (1.1), we assume

$$c_\theta = \pi_\theta c_1, \text{ where } c_1 < c_0. \quad (1.48)$$

The assumption states that set-up capital required for a robust technology is larger than the capital required for risky technologies. Moreover, the smaller the measure π_θ of the state under which a set-up technology works, the lower the required set-up capital.²¹ Robust set-up technologies are financed by loans, whereas the risky set-up techniques are financed by state-contingent securities.

Let K_0 be the aggregate set-up capital for robust technologies and denote by $K_\theta, \theta \in \Theta$, the aggregate set-up capital for specialized risky technologies. Then, the number of firms which can be set up is $M_0 = \frac{K_0}{c_0}$ and $M_\theta = \frac{K_\theta}{c_\theta}$, respectively. In a closed economy, capital markets are cleared if

$$K_0 = D, \quad K_\theta = F_\theta = \pi_\theta F.$$

²⁰Note that $\log \frac{e - \bar{e}}{P} = \log(e - \bar{e}) - \log P$ so that the P -levels add to $\mathbb{E}U$ a constant.

²¹See Falkinger (2014) for a more detailed discussion of the relationship between specialization and risk. There, technologies are more productive the more narrowly they are targeted to a specific environment. At the same time, they are more risky because the realization of the specific environment is less likely. Here this idea is applied to set-up costs rather than productivity.

Hence, we have for to total number of firms

$$M = \begin{cases} \frac{D}{c_0} + \frac{F}{c_1} \equiv M_\Theta, & \text{if } \theta \in \Theta, \\ \frac{D}{c_0} \equiv M_{\bar{\Theta}} & \text{otherwise.} \end{cases}$$

After firms being set up, their operating profits earned under mark-up prices (1.47) are

$$\Pi = (p_x - c)X = \frac{X}{\sigma},$$

where $p_x = 1$, which implies $c = \frac{\sigma-1}{\sigma}$, has been used. Since firms are symmetric, aggregated operating profits are distributed uniformly across firms so that operating profit per firm is:

$$\frac{\Pi_m}{M_m} = \frac{X}{\sigma M_m}, \quad m \in \{\Theta, \bar{\Theta}\}.$$

The returns on one unit of set up capital are therefore

$$r_m = \frac{X}{c_0 \sigma M_m}, \quad m \in \{\Theta, \bar{\Theta}\}$$

$$R_\theta = \frac{X}{c_\theta \sigma M_\Theta}, \quad R = \frac{X}{c_1 \sigma M_\Theta}$$

for safe and risky investments, respectively. ($\pi_\theta R_\theta$ reduces to R because of assumption $c_\theta = \pi_\theta c_1$.) Since the number of firms is different in Θ and $\bar{\Theta}$, aggregate operating profits have to be shared among more or less firms so that the return on robust investments is m -dependent. The relative rate of return, $\frac{r_\Theta}{R_\theta}$, however, is uniquely determined by the relative set-up requirements of specialized risky technologies compared to the robust technology. We have

$$\rho = \frac{c_1}{c_0}.$$

For the portfolio choice derived in Section 1.4 almost only the relative rate ρ matters. The exception is $\frac{\bar{e}_1}{r_m}$, since future subsistence can only be financed by deposits.²² This means, we have to restrict the analysis of the paper to $\bar{e}_1 = 0$, or we reconcile the fluctuation of the earnings of robust firms with a safe return on deposits by assuming that firms hold buffers and distribute the expected profit per firm $\bar{\pi} \equiv [\frac{\mu}{M_\Theta} + \frac{1-\mu}{M_{\bar{\Theta}}}] \frac{X}{\sigma}$ to

²²Formally the derivation of the portfolio choice presented in the appendix has to be adapted to account for m -dependent pay-offs in the budget constraints. For $\bar{e}_1 = 0$, return $r_{\bar{\Theta}}$ becomes irrelevant under the log-specification and the analysis remains valid – with $\rho = \frac{r_\Theta}{R_\Theta}$.

the investors.

For the general equilibrium analysis, a further caveat is in order. Under the presented extension, return r (even if smoothed by the buffer) is endogenous. It depends on M and X , which are determined by saving behavior and resource allocation, respectively. Thus, in the general equilibrium, a further feedback loop is to be considered. We did not account for such feedbacks in Section 1.6, since in the baseline return r is exogenously given by the constant productivity of capital. For $\bar{e}_1 = 0$, however, the presented analysis remains fully valid also with set-up capital of firms, since r matters only through the term $\frac{\bar{e}_1}{r}$. However, what one loses by setting $\bar{e}_1 = 0$ is the income effect on structural change within the financial sector. For the income effect on the skill premium and the structural change between goods and financial sector subsistence level $\bar{e}_0 > 0$ is relevant, which poses no problem in the extension considered here.

1.8 Robustness

To account for relative price effects within the financial sector, we skip now Assumption 1.2 and impose the following assumption instead.

Assumption 1.2' $\alpha_x = \alpha_{z_1} < \alpha_{z_2}$.

Then, according to (1.12),

$$p_{z_1} = \frac{A_x}{A_{z_1}}$$

and thus: $\bar{y} = \bar{e}_0 + \frac{(1 + \frac{A_x}{A_{z_1}})\bar{e}_1}{r}$.

Moreover, the terms $a_x^l X + a_{z_1}^l Z_1$, $l \in \{H, L\}$, in system (1.13) reduce to

$$X^+ \frac{1}{A_x \kappa_x^{\alpha_x}} \text{ and } X^+ \frac{\kappa_x^{(1-\alpha_x)}}{A_x}, \quad X^+ \equiv X + \frac{A_x}{A_{z_1}} Z_1,$$

respectively. Using this when solving (1.13), we obtain

$$X^+ = \frac{b_L \bar{L}}{a_x^L} \frac{\kappa_{z_2} - k}{\kappa_{z_2} - \kappa_x}, \quad Z_2 = \frac{b_L \bar{L}}{a_{z_2}^L} \frac{k - \kappa_x}{\kappa_{z_2} - \kappa_x} \quad (1.49)$$

and

$$\frac{Z_2}{X^+} = \frac{A_{z_2}}{A_x} \tilde{\gamma}(\omega, k), \quad \tilde{\gamma}(\omega, k) \equiv \frac{\kappa_{z_2}^{\alpha_{z_2}}}{\kappa_x^{\alpha_x}} \frac{k - \kappa_x}{\kappa_{z_2} - k}, \quad (1.50)$$

where the signs for the partial derivatives of $\tilde{\gamma}$ follow from the Rybczynski analysis.

Substituting $A_{z_2} \kappa_{z_2}^{\alpha_{z_2}}$ for $\frac{1}{a_{z_2}^L}$ in the second equation of (1.49) and using (1.9), we have for the Z_2 -supply:

$$Z_2^S = A_{z_2} b_L \bar{L} \frac{\gamma_{z_2}^{\alpha_{z_2}}}{\gamma_{z_2} - \gamma_x} g(\omega, k), \quad g(\omega, k) \equiv \omega^{-\alpha_{z_2}} (k\omega - \gamma_x). \quad (1.51)$$

This coincides with (1.40) – with Z_2 instead of Z – so that Fact 1.6 remains valid under the alternative specification and applies to Z_2 -supply.

Z_2 -demand is given by

$$Z_2^D = F = s_f \frac{\delta}{1 + \delta} \frac{\bar{w} - \bar{y}}{1 + p_{z_2}} N = \frac{\mu - \rho p}{1 - \rho p} \frac{\delta}{1 + \delta} \tilde{\eta} N \quad (1.52)$$

with $\tilde{\eta} \equiv \frac{\bar{w} - \bar{y}}{1 + p_{z_2}}$ and $p = \frac{1 + p_{z_2}}{1 + p_{z_1}}$. In an analogous way to Lemma 1.1 and Fact 1.7, one establishes that the income effect (i.e., $\tilde{\eta}$ -part in Z_2^D) has a U-shaped form.²³ Further, s_f is decreasing in ω since $\frac{\partial p}{\partial \omega} > 0$ (according to (1.12)). Because of the relative price effect p , which now is at work within the finance sector, the demand for risky assets is substituted by demand for safe assets if the relative price of services for securities rises. For low values of the skill premium, we are on the downward sloping branch of the $\tilde{\eta}$ -curve so that income and substitution effect go in the same direction. In the upward sloping part of $\tilde{\eta}$, the negative substitution effect is opposed by a positive income effect so that the total effect of ω on Z_2^D depends on the relative importance of the two effects. Numerical simulation shows that the substitution effect is large if the price p_{z_2} is high and the income effect is stronger if subsistence expenditures are larger. For a high level of price p_{z_2} (based on (1.12) this means, for example, a low A_{z_2}) and low subsistence levels (such that \bar{y} is close to zero) the substitution effect dominates. In this case $\frac{\partial Z_2^D}{\partial \omega} < 0$. However, for low levels of price p_{z_2} and large subsistence levels the income effect dominates. For this case, (1.51) and (1.52) give us the same picture as in Figure 1.1. Proposition 1.4 remains valid in both cases.

For results corresponding to Proposition 1.5, we have to write the excess demand function $Z_2^D - Z_2^S$ explicitly in terms of parameters. Using $W = b_L \bar{L} A_x \Gamma_x \omega^{-\alpha_x} (1 + \omega k)$ and $p_{z_2} = \frac{A_x \Gamma_x}{A_{z_2} \Gamma_{z_2}} \omega^{\alpha_{z_2} - \alpha_x}$ in (1.52), we can rewrite the equilibrium condition $Z_2^D - Z_2^S = 0$

²³The only thing that changes is that now we have $\frac{\bar{y}}{1 + p_{z_2}}$ with \bar{y} constant instead of $\frac{\bar{y}}{1 + p_{z_1}} = \frac{\bar{e}_0}{1 + p_{z_1}} - \frac{\bar{e}_1}{r}$. Thus, apart from subscript z_2 instead of $z(=z_1=z_2)$ in the modified proof we have \bar{y} instead of \bar{e}_0 and no negative term $-\frac{\bar{e}_1}{r}$.

in the form:

$$\begin{aligned} & \frac{\mu - \rho \frac{1 + \frac{A_x \Gamma_x}{A_{z_2} \Gamma_{z_2}} \omega^{\alpha_{z_2} - \alpha_x}}{1 + p_{z_1}}}{1 - \rho \frac{1 + \frac{A_x \Gamma_x}{A_{z_2} \Gamma_{z_2}} \omega^{\alpha_{z_2} - \alpha_x}}{1 + p_{z_1}}} \frac{\delta}{1 + \delta} \frac{\Gamma_x \omega^{-\alpha_x} (1 + \omega k) - \frac{N}{b_L \bar{L} A_x} \bar{y}}{1 + \frac{A_x \Gamma_x}{A_{z_2} \Gamma_{z_2}} \omega^{\alpha_{z_2} - \alpha_x}} - \frac{A_{z_2}}{A_x} \frac{\gamma_{z_2}^{\alpha_{z_2}}}{\gamma_{z_2} - \gamma_x} \omega^{-\alpha_{z_2}} (k\omega - \gamma_x) \\ & \equiv D \left[\omega \left| \frac{A_{z_2}}{A_x}, k, \frac{A_x b_L \bar{L}}{N}, \bar{y}, \mu, \rho, \delta \right| \right] = 0. \end{aligned}$$

Hence, an increase of $\frac{A_x b_L \bar{L}}{N}$ always leads to a rise in the equilibrium skill premium. Under Assumption 1.2, this was only the case if present subsistence expenditure dominates futures subsistence requirements (Proposition 1.5). Moreover, a decline in subsistence requirements \bar{y} has unambiguously a positive impact on the equilibrium skill premium - regardless of whether the decline in \bar{y} is caused by a decline in \bar{e}_0 or \bar{e}_1 .

In contrast to the benchmark analysis, the equilibrium skill premium is now also affected by changes in μ and ρ . Finally, a rise in δ has now an unambiguously positive effect on ω^* . (In the benchmark analysis the role of δ was ambiguous.) The following proposition summarizes the comparative-static effects on the equilibrium skill premium under Assumption 1.2'.

Proposition 1.5'. *If Assumption 1.2 is replaced by Assumption 1.2', then:*

- a) *For $\bar{y} > 0$, a rise in $\frac{A_x b_L \bar{L}}{N}$ (caused by uniform technical progress or education and biased progress) raises the equilibrium skill premium. A decline of total subsistence requirements \bar{y} (wherever they come from) have the same effect.*
- b) *Financial innovation (a rise in μ) or increased attractiveness of risky investments (a decline of ρ) raise the equilibrium skill premium. A lower discount on the future (a rise of δ) has the same effect. These effects also hold if $\bar{y} = 0$.*

Proof. Main text. □

As a consequence of (1.50), Proposition 1.6 remains valid if applied to the structure between new finance on the one side and production cum traditional finance on the other side. We have

Proposition 1.6'. *At given $\frac{A_z}{A_x}$, k , any change in other exogenous fundamentals which raises the skill premium leads to structural change from production and traditional finance (X^+) towards new finance (Z_2).*

Proof. Equation (1.50) and $\frac{\partial p_{z_2}}{\partial \omega} > 0$. □

Finally, the ratio of value added in financial subsector Z_2 to value added in subsector Z_1 is as in (1.33)

$$\frac{p_{z_2}F}{p_{z_1}D} = \frac{s_f \bar{\eta}}{s_d \bar{\eta} + \frac{1+\delta}{\delta} \frac{e_1}{r}} \frac{p_{z_2}}{1+p_{z_2}} \frac{1+p_{z_1}}{p_{z_1}}. \quad (1.53)$$

Since p_{z_1} and \bar{y} are constant, $\frac{\partial \bar{\omega}}{\partial \omega} > 0$ immediately implies $\frac{\partial \bar{\eta}}{\partial \omega} > 0$. Hence, for $\bar{e}_1 > 0$, the income effect unambiguously leads to structural change from Z_1 to Z_2 if the skill premium rises. If $\bar{e}_1 = 0$, no such income effect is at work; yet the relative price effect remains. For the relative price effect, we only have to consider p_{z_2} because p_{z_1} is constant. Price p_{z_2} affects the value added structure within finance through two channels: On the one side, there is the direct effect shown explicitly in (1.53). Since $\frac{\partial p_{z_2}}{\partial \omega} > 0$, this channel tends to increase the share of new finance. On the other side, however, there is the negative substitution effect in the demand for financial services ($\frac{\partial s_f}{\partial p} < 0$ and $\frac{\partial s_d}{\partial p} > 0$) which drives the sectoral structure within finance from Z_2 towards Z_1 . Due to this ambiguous role of the relative price effect under the alternative specification, within structural change from Z_1 to Z_2 is more difficult to model than it was in the benchmark. For high levels of price p_{z_2} and low subsistence expenditures the substitution effect dominates. Then, the presented model cannot predict a co-movement of ω and the within structural change from Z_1 to Z_2 . In the other case, however, Proposition (1.7) applies.

1.9 Empirical evidence and numerical exercises

In this section we first provide empirical evidence for the two-fold structural change and wage inequality and then we carry out numerical exercises to illustrate how the presented model can replicate the observed changes.

1.9.1 Empirics

1.9.1.1 Data description

We use data from the Current Population Survey (March CPS) for the survey years 1980-2013 from IPUMS-CPS by King et al. (2010).²⁴ This data set allows us to split the sampled population (weighted with the sampling weight) into our three sectors and two skill levels: The X -sector consists of all sectors of the U.S. economy except finance. The finance sector is finance and insurance without real estate.²⁵ “Traditional finance” Z_1 includes banking, credit agencies and insurance. “New finance” Z_2 is security and

²⁴Survey years 1980-2013 represent years 1979-2012 because households are surveyed about last year’s job. This means whenever we talk about a year the data considered represent the situation a year before.

²⁵This corresponds to the standard classification as in Philippon and Reshef (2007, 2012).

commodity brokerage and investment companies. We define a worker (if worked positive weeks last year) to be high-skilled if she/he holds a college degree (four-year college) or more. \bar{H}_j is the number of high-skilled workers in sector $j \in \{x, z_1, z_2\}$ and \bar{L}_j is the number of low-skilled workers in sector $j \in \{x, z_1, z_2\}$. For each skill level, we calculate for the three sectors the average yearly hours worked last year (i.e., h_j^l , $j \in \{x, z_1, z_2\}$, $l \in \{H, L\}$) and the respective average hourly real wages (i.e., w_j^l , $j \in \{x, z_1, z_2\}$, $l \in \{H, L\}$).²⁶

In our data analysis we use “actual” and “normalized” numbers for employment and wage levels. The “actual” numbers use the observed sector- and skill-specific average yearly hours worked and the respective average hourly wage. The “normalized” numbers are calculated all with the same basis of hours worked and hourly wage (i.e., the ones from the X -sector).²⁷ This normalization allows us to separate the effects we can identify in the theoretical, frictionless model from two frictions which are observed in reality: (i) Low- and high-skilled Z -workers work more hours per year than low- and high-skilled X -workers. More precisely, for the U.S. over the last decades on average a Z -worker has worked about 9% more than a X -worker. (ii) There is the finance premium on hourly wages for low- and high-skilled Z -workers.²⁸ CPS data show that the finance premium increased over time and differs for the two subsectors: In Z_1 workers earn about 15% more than in the X -sector, in Z_2 it is even 50%.

The two sectoral structure-figures below show black and gray lines: Gray lines correspond to the “actual” numbers. Black lines correspond to the “normalized” ones.

²⁶We use worker’s total pre-tax wage and salary income to calculate average hourly real wages (nominal values are adjusted by using the CPI-U adjustment factor to 1999 dollars (i.e., for the base survey year 2000)). There are two issues related to this: First, the CPS top-codes high wage incomes for reasons of confidentiality. This leads to an underestimation of wages in general and especially in the finance sector: Over all our survey years around 0.8% of workers in the X -sector are top-coded whereas in the Z_1 -sector it affects around 1.6% of the workers and in the Z_2 -sector even 7.6%. To dampen the bias in high wages we multiply top-coded incomes for survey years 1980-1995 by 1.5; a standard factor used in literature (as it is described in Philippon and Reshef (2007, 2012)). From year 1996 on, top-coded wages are categorized into groups with different mean incomes by the CPS and thus the aggregate and the average wage income are uninfluenced by the top-coding. Note that our results are not very sensitive with respect to the multiplication factor (e.g., as compared to $\omega = 1.62$, $\Psi = 5.08\%$ and $\Phi = 13.99\%$ in Table 1.2 resulting from multiplication factor 1.5, using a factor of 1.75 as in Philippon and Reshef (2007, 2012) would result in $\omega = 1.63$, $\Psi = 5.09\%$ and $\Phi = 14.02$). Second, total wage income consists of both wage income from longest job last year and wage income from other work. We cannot allocate these two incomes to different industries nor to the respective hours worked. Thus, we allocate the total wage income to one industry and account for it all hours worked. If one assumes that the switch of job occurs equally likely between the three sectors, this does not bias the results. Furthermore, only about one fifth of all workers (in all three sectors) is affected by this and of those who are affected not even a fourth of total income is coming from other work.

²⁷Since the skill premium is approximately identical in all three sectors in the U.S. the skill intensities in the sectors need not be “normalized”. They already correspond to the frictionless numbers.

²⁸See Célérier and Vallée (2015) or Philippon and Reshef (2007, 2012) for a detailed empirical discussion of the finance premium.

1.9.1.2 Empirical trends

As is described in the introduction and picked up in the model, financialization has several aspects: On the one hand, the weight of the financial sector relative to non-financial business has increased; this is structural change towards finance. On the other hand, the type of financial products and services has changed; this is structural change within finance. The next two figures show the two-fold structural change.

Figure 1.2 shows the ratio of the total finance sector (Z -sectors) compared to the non-finance economy (X -sector) for the U.S. based on the CPS data.

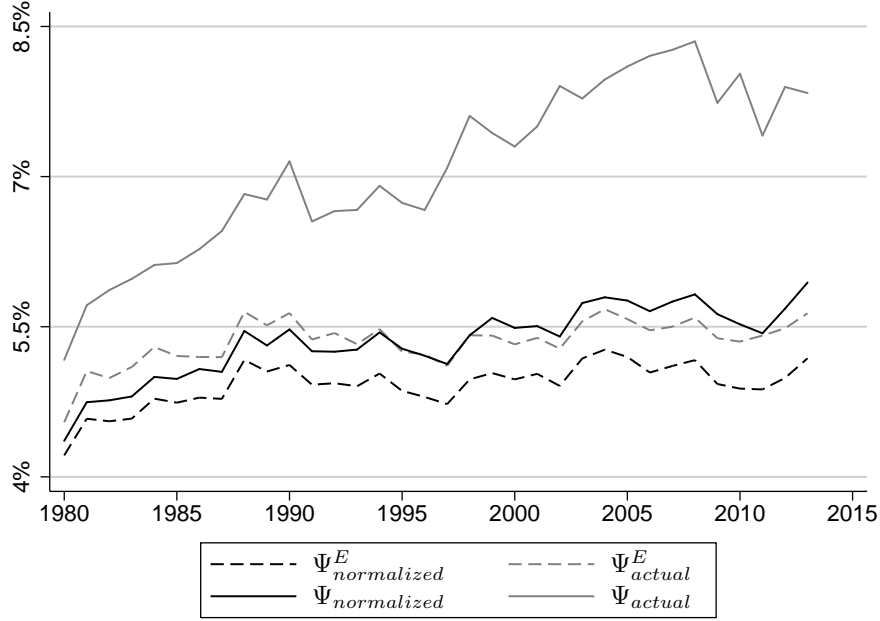


Figure 1.2: Employment ratio and wage sum ratio of the financial sector

Notes: Ψ^E measures the employment ratio (in terms of total hours worked) of finance (including insurance) compared to the rest of the U.S. economy. Ψ measures the ratio of the total wage sum in finance vs. the rest of the U.S. economy. “Actual” uses the observed sector-specific hours worked and hourly wages (for low-and high-skilled), whereas “normalized” uses the X -sector hours worked and hourly wages (for low-and high-skilled). Survey years from 1980-2013. Source: Own calculations based on CPS.

On the one hand, the figure shows that finance has attracted new employment. The employment ratio (in terms of total hours worked) of the financial sector, defined by $\Psi_{actual}^E \equiv \frac{h_{z1}^H \bar{H}_{z1} + h_{z2}^H \bar{H}_{z2} + h_{z1}^L \bar{L}_{z1} + h_{z2}^L \bar{L}_{z2}}{h_x^H \bar{H}_x + h_x^L \bar{L}_x}$, increased from 4.5% in 1980 to 5.6% in 2013. The respective “normalized” ratio $\Psi_{normalized}^E \equiv \frac{h_x^H \bar{H}_{z1} + h_x^H \bar{H}_{z2} + h_x^L \bar{L}_{z1} + h_x^L \bar{L}_{z2}}{h_x^H \bar{H}_x + h_x^L \bar{L}_x}$ rose from 4.2% in 1980 to 5.2% in 2013. On the other hand, the figure illustrates the structural change towards the financial sector in terms of a growing wage sum ratio of finance. The wage sum ratio of the financial sector, defined as $\Psi_{actual} \equiv \frac{w_{z1}^H h_{z1}^H \bar{H}_{z1} + w_{z2}^H h_{z2}^H \bar{H}_{z2} + w_{z1}^L h_{z1}^L \bar{L}_{z1} + w_{z2}^L h_{z2}^L \bar{L}_{z2}}{w_x^H h_x^H \bar{H}_x + w_x^L h_x^L \bar{L}_x}$, increased by 50% from about 5.2% in 1980 to 7.8% in 2013. The respective “normalized” ratio $\Psi_{normalized} \equiv \frac{w_x^H h_x^H \bar{H}_{z1} + w_x^H h_x^H \bar{H}_{z2} + w_x^L h_x^L \bar{L}_{z1} + w_x^L h_x^L \bar{L}_{z2}}{w_x^H h_x^H \bar{H}_x + w_x^L h_x^L \bar{L}_x}$ rose by 34% from 4.4% in 1980

to 5.9% in 2013. The difference between the employment (E) ratio and the wage sum ratio is the result of different skill-intensities in the different sectors. By comparing the “normalized” black with the “actual” gray lines one sees a large difference between the two ratios of the wage sum: More than half of the increase in the ratio of the wage sum is the result of the frictions (i) (more hours) and (ii) (finance premium). Yet, as the black line shows, there is still structural change towards finance if one controls for the two frictions. Comparison of the two black lines shows that the difference between the employment ratio and the wage sum ratio increased over time.

We observe a similar pattern for the within finance sectoral structure by splitting total finance up into the two subsectors Z_1 and Z_2 . Figure 1.3 shows the employment ratio and the wage sum ratio of finance subsector Z_2 compared to the subsector Z_1 for the U.S. since the 1980s based on the CPS data set.

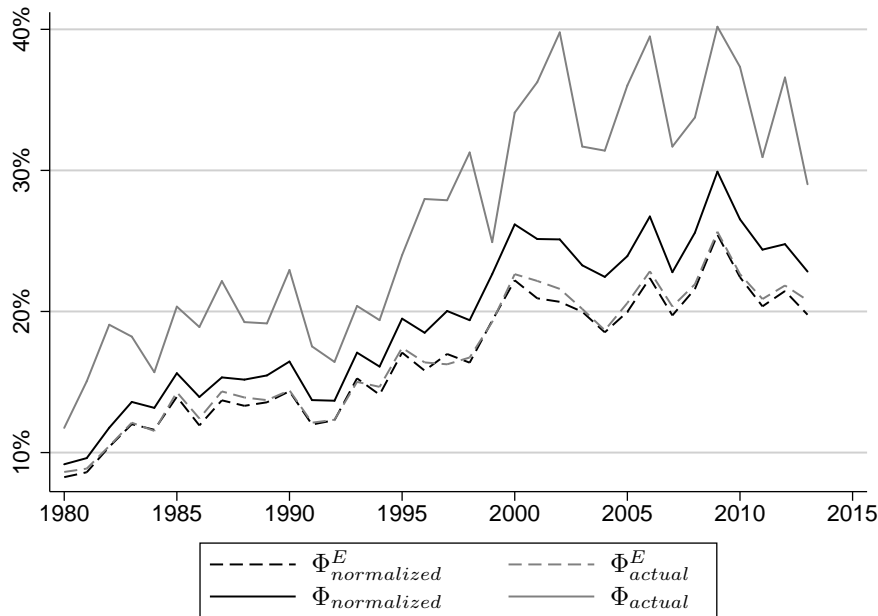


Figure 1.3: Employment ratio and wage sum ratio within the financial sector

Notes: Φ^E measures the employment ratio (in terms of total hours worked) of “new finance” compared to “traditional finance”. Φ measures the ratio of the total wage sum in “new finance” vs. “traditional finance”. “Actual” uses the sector-specific hours worked and hourly wages (for low- and high-skilled), whereas “normalized” uses the X -sector hours worked and hourly wages (for low- and high-skilled). Survey years from 1980–2013. Source: Own calculations based on CPS.

“New finance” (subsector Z_2) grew strongly independent of the measure we use: The within employment ratio (in terms of total hours worked) of finance subsector Z_2 , $\Phi_{actual}^E \equiv \frac{h_{z_2}^H \bar{H}_{z_2} + h_{z_2}^L \bar{L}_{z_2}}{h_{z_1}^H \bar{H}_{z_1} + h_{z_1}^L \bar{L}_{z_1}}$, more than doubled from about 8.6% in 1980 to 20.8% in 2013. The respective “normalized” ratio $\Phi_{normalized}^E \equiv \frac{h_x^H \bar{H}_{z_2} + h_x^L \bar{L}_{z_2}}{h_x^H \bar{H}_{z_1} + h_x^L \bar{L}_{z_1}}$ is very similar with a rise from 8.3% in 1980 to 19.8% in 2013. The within finance wage sum ratio, defined by $\Phi_{actual} \equiv$

$\frac{w_{z2}^H h_{z2}^H \bar{H}_{z2} + w_{z2}^L h_{z2}^L \bar{L}_{z2}}{w_{z1}^H h_{z1}^H \bar{H}_{z1} + w_{z1}^L h_{z1}^L \bar{L}_{z1}}$, increased dramatically from 11.8% in 1980 to 29% in 2013 peaking in survey 2009 at 40.2%. The respective “normalized” ratio $\Phi_{normalized} \equiv \frac{w_x^H h_x^H \bar{H}_{z2} + w_x^L h_x^L \bar{L}_{z2}}{w_x^H h_x^H \bar{H}_{z1} + w_x^L h_x^L \bar{L}_{z1}}$ rose from 9.2% in 1980 to 22.8% in 2013 with a peak in survey year 2009 of 29.9%. Hence, about two-thirds of the actual rise in the wage ratio of “new finance” cannot be assigned to frictions: Structural change is also observed in the “normalized” data. The rest of the rise comes from friction (ii) (finance premium), which is particularly strong in the finance subsector Z_2 .

As argued in the introduction financialization (with the two-fold structural change) and inequality are two closely related topics. Figure 1.4 shows the development of the “normalized” skill premium calculated by $\omega = \frac{w_x^H}{w_x^L}$ for the U.S. since 1980 based on the CPS data.²⁹

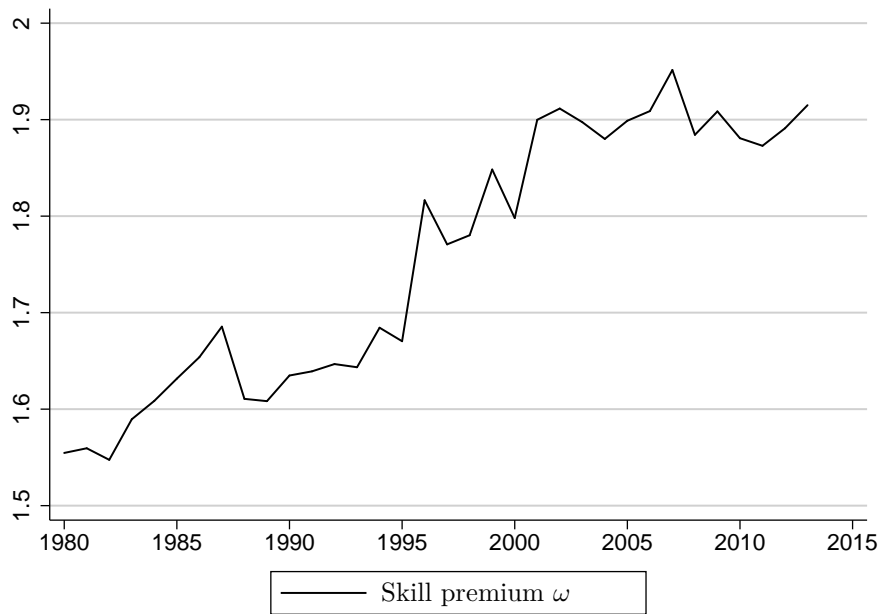


Figure 1.4: Skill premium

Notes: ω measures the “normalized” skill premium (i.e., hourly wage of high-skilled labor in X -sector divided by hourly-wage of low-skilled labor in X -sector). Survey years from 1980-2013. Source: Own calculations based on CPS.

It increased from 1.55 in 1980 to 1.91 in 2013. This time trend in ω illustrates that wage inequality increased over time. Nowadays high-skilled workers earn nearly double as much as low-skilled workers per hour. If one accounts additionally for the fact that high-skilled workers work more hours, the income inequality is even larger (e.g., 2.19 in 2013).

²⁹Interestingly, the skill premium in the U.S. is about the same in the three sectors because both low- and high-skilled workers in the financial industry earn a similar relative finance premium.

1.9.2 Numerics

In this section we implement our theoretical model quantitatively and use it for several numerical exercises. These illustrate possible drivers of the empirical developments presented in the Figure 1.2-1.4. First, we calibrate our model for the average value of the early survey years 1980-1994. Second, this calibrated model is used for comparative-static analysis. We introduce (i) *ceteris paribus* shocks and (ii) simultaneous shocks to illustrate how the channels analyzed in our model can predict the situation observed in later years (average values of later survey years 1995-2009).

1.9.2.1 Calibration

We calibrate our model such that it fits the data for the average of the survey years 1980-1994 (i.e., years 1979-1993). Table 1.1 gives the exogenous parameters used.

Table 1.1: Parameters survey years 1980-1994

Parameter	Data	Source	Description
\bar{L}	99.2m	CPS	# Low-skilled employees
\bar{H}	26.5m	CPS	# High-skilled employees
h^L	1639.4	CPS	Yearly hours of low-skilled
h^H	1982.6	CPS	Yearly hours of high-skilled
α_x	0.34	CPS	Output ela. of high-skilled in X
α_{z_1}	0.42	CPS	Output ela. of high-skilled in Z_1
α_{z_2}	0.68	CPS	Output ela. of high-skilled in Z_2
A_x	26.53	CPS	Technology level in X
PT_{65}	\$ 11,204	U.S. Bureau of the Census	Real poverty threshold <65
PT^{65}	\$ 10,076	U.S. Bureau of the Census	Real poverty threshold >65
$LEratio$	4.66	LE from World Bank	Old-age ratio
r^f	0.0368	Federal Reserve Bank of St.Louis	Real effective federal funds rate
A_{z_1}	116	Model calibration	Technology level in Z_1
A_{z_2}	165	Model calibration	Technology level in Z_2
δ	0.385	Model calibration	Discount rate
μ	0.740	Model calibration	Certainty measure

Notes: The table shows the averaged values for the time range of survey years $t \in \{1980, \dots, 1994\}$. Averages of $\alpha_{j,t} = \frac{\kappa_{j,t}\omega_{j,t}}{1+\kappa_{j,t}\omega_{j,t}}$ with $\kappa_{j,t} = \frac{h_{j,t}^H \bar{H}_{j,t}}{h_{j,t}^L L_{j,t}}$ and $\omega_{j,t} = \frac{w_{j,t}^H}{w_{j,t}^L}$, $j \in \{x, z_1, z_2\}$, $h_t^H = h_{x,t}^H$ and $h_t^L = h_{x,t}^L$. $A_{x,t} = \frac{w_{x,t}^L}{\Gamma_{x,t}\omega_{x,t}^{1-\alpha_{x,t}}}$ with $\Gamma_{x,t} = \alpha_{x,t}^{\alpha_{x,t}}(1 - \alpha_{x,t})^{1-\alpha_{x,t}}$. PT is the average, real poverty threshold of a two-people household (nominal values are adjusted by using the CPI-U adjustment factor to 1999 dollars (i.e., for the base survey year 2000) from CPS with PT_{65} denoting the relevant value for households younger than 65 and PT^{65} denoting the value relevant for older ones. $LEratio$ is the average ratio of working-time to retirement: $(65 - 20)/(LE_t - 65)$, where LE_t denotes life expectancy in year t ; 65 is the retirement age and 20 is the assumed start of the working-life. r^f is the average, real effective federal funds rate (effective federal funds rate adjusted with the CPI-U adjustment factor from CPS). See bibliography for details on data sources.

Exogenous values from data are used for labor endowments $\bar{L}, \bar{H}, h^L, h^H$, output elasticities α_j , technology in the X -sector A_x , poverty thresholds (PT_{65} for young and PT^{65} for old households) and interest rate r^f as summarized in Table 1.1. For the subsistence levels we assume that each worker must cover over the life cycle half of a two-people household's poverty threshold. Further, we account for the fact that during 1980-1994 the ratio of working-time to retirement was $LEratio = 4.66$ (i.e., we divide the poverty threshold of old households by 4.66). Hence, $\bar{e}_0 = PT_{65}/2$ and $\bar{e}_1 = PT^{65}/2/4.66$. The real safe return is $r = 1 + r^f$ with r^f being the real effective federal funds rate and the risky return is such that the risk premium is four percentage points (i.e., $R = (r + 0.04)/\mu$). We measure the efficiency units from the model by $b_l = h^l$, $l \in \{H, L\}$, where h^l are hours worked.

The other parameters (productivities in the finance sectors A_{z_1} and A_{z_2} , discount factor δ and completeness measure μ) are calibrated internally by targeting wage inequality ω , “normalized” ratios for the sectoral structure Ψ and Φ of the U.S. economy and the gross saving rate in the U.S. for the average of the survey years 1980-1994. The targeted values are shown in Table 1.2. More specifically, we solve the model numerically for possible parameter combinations of A_{z_1} , A_{z_2} , δ and μ and grid-search for the combination (see Table 1.1 for calibrated values) which minimizes the sum of the squared relative distances of the four model values from the corresponding data targets.³⁰ The comparison of the four model values generated by our calibrated model with the data outcomes is given in Table 1.2: The calibrated model fits the targets fairly well.

Table 1.2: Targets

Variables	Model	Data	Source	Description
ω^*	1.63	1.62	CPS	Skill premium
Ψ	5.08%	5.08%	CPS	Between sectoral structure
Φ	13.99%	13.99%	CPS	Within sectoral structure
saving rate	20.32%	20.30%	World Bank	Aggregate savings

Notes: ω^* is the equilibrium skill premium (per hour worked). Ψ corresponds to $\frac{p_{z_1}D + p_{z_2}F}{X}$ in the model and to $\Psi_{normalized}$ in the data. Φ corresponds to $\frac{p_{z_1}D}{p_{z_2}F}$ in the model and to $\Phi_{normalized}$ in the data. The saving rate is $(D + F)/W$ in the model and the share of aggregate savings in gross national income in the data, where aggregate savings (gross savings) is gross national income less total consumption, plus net transfers. See bibliography for details on data sources.

³⁰For solving the model numerically, we use the demand functions in the goods and financial services markets to obtain the equilibrium values of X -, Z_1 - and Z_2 as functions of ω (and exogenous parameters). Substituting these functions for X -, Z_1 - and Z_2 in one of the labor market clearing conditions, we can solve for the equilibrium skill premium ω^* . (Then, at ω^* , the other labor market is also cleared.) From ω^* follow factor prices and prices of financial services, output levels and employment in the three sectors and the sectoral structure of the economy in a straightforward way.

Further, the other equilibrium values following from the model are also very similar to the values observed in the CPS data (given in brackets). Hourly wages in our model are $w^H = \$19.3$ (\$19.3), $w^L = \$11.8$ (\$11.9) and the resulting prices are $p_{z_1} = 0.25$, $p_{z_2} = 0.19$.³¹ Labor employments in total hours are $H_x = 49215\text{m}$ (49215m), $L_x = 156043\text{m}$ (156287m), $H_{z_1} = 2710\text{m}$ (2794m), $L_{z_1} = 6113\text{m}$ (6022m), $H_{z_2} = 614\text{m}$ (635m), $L_{z_2} = 472\text{m}$ (468m). For the skill intensities we get $\kappa_x = 0.32 < \kappa_{z_1} = 0.44 < \kappa_{z_2} = 1.30$ ($\kappa_x = 0.31 < \kappa_{z_1} = 0.43 < \kappa_{z_2} = 1.30$), which shows that the two finance subsectors are more skill intensive than the rest of the economy. These numbers suggest that the calibrated model matches the U.S. economy in the survey period 1980-1994 fairly well.

1.9.2.2 Numerical exercises

We show now how our calibrated model can predict the twofold structural change and the rising wage inequality between survey period 1980-1994 and survey period 1995-2009 as seen in the Figure 1.2-1.4. To do so, we look at the predictions of our calibrated model if shocked by exogenous changes. Thereby, we apply the changes in the exogenous parameters of our model as observed in data. In other words, we use as shocks the average values of \bar{L} , \bar{H} , h^L , h^H , α_x , α_{z_1} , α_{z_2} , A_x , PT_{65} , PT^{65} , $LEratio$ and r^f for the time span of the survey years 1995-2009 instead of the ones for the time span of the survey years 1980-1994.³² In addition, we also consider shocks on the internally calibrated parameter A_{z_1} , A_{z_2} , δ and μ .

As a first exercise, we introduce ceteris paribus shocks. This means that we separately apply each of the changes listed in Table 1.3. We apply observed changes for the exogenous parameters and potential changes for the internally calibrated parameters. The qualitative effects of such ceteris paribus changes on the skill premium ω , on the between sectoral structure Ψ and the within structure Φ are summarized in Table 1.3. These comparative-static effects can be interpreted as follows: Uniform productivity progress A_j means that the productivities in all three sectors $j \in \{X, Z_1, Z_2\}$ grow at the same rate (i.e., $A_{z_i}^1 = g_x A_{z_i}^0$, $i \in \{1, 2\}$, where $g_x = A_x^1/A_x^0$ is given by the observed average values of A_x^0 from survey years 1880-1994 and of A_x^1 from survey years 1995-2009). Consistent with Proposition 1.5-1.7 such a uniform productivity progress leads to an increase in the skill premium as well as to the twofold structural change. This is due to the income effect arising through the subsistence requirements $\bar{e}_0 > 0$ and $\bar{e}_1 > 0$. Sector-biased technical

³¹The magnitude of the financial services prices could be interpreted in the following way: A household has to pay the unit costs of financial intermediation, estimated by Philippon (2015) to be 0.015-0.02, during all his/hers “capital-accumulation” years (i.e., 15-times from 1980-1994 to 1995-2009).

³²See Table A.1 in Appendix A.4 for data of the average values for survey years 1995-2009 of \bar{L} , \bar{H} , h^L , h^H , α_x , α_{z_1} , α_{z_2} , A_x , PT_{65} , PT^{65} , $LEratio$ and r^f . For R we use again a constant risk premium of four percentage points.

Table 1.3: Comparative statics

	ω	Ψ	Φ
Uniform productivity progress A_j (income effect)	+	+	+
X -biased technical change A_x	+	+	+
Z_1 -biased technical change A_{z_1}	-	-	-
Z_2 -biased technical change A_{z_2}	-	-	+
Skill-biased technical change $\alpha_x, \alpha_{z_1}, \alpha_{z_2}$	+	-	-
Higher subsistence requirement young \bar{e}_0	-	-	-
Higher subsistence requirement old \bar{e}_1	+	+	-
Increased skill supply k	-	-	+
Lower safe return r ($\frac{\bar{e}_1}{r}$ -channel)	+	+	-
Lower relative return ρ	+	-	+
More completeness μ	+	-	+
Fall in δ	-	-	-

Notes: + is a positive comparative-static effect. - is a negative comparative-static effect.

change means that only the respective sector grows, while the other two productivity levels are kept constant (as growth rate we use always the observable rate g_x). The comparative-static effects of such a ceteris paribus shock are a combination of income and substitution effects. (Sector-specific) skill-biased technical change α_j , as observed in the data for $j \in \{X, Z_1, Z_2\}$, induces clearly an increase of the skill premium. Higher subsistence requirements (mainly a higher \bar{e}_1 because of aging households) lead to similar effects as predicted in Proposition 1.5-1.7. An increase in skill supply $k = \frac{\bar{H}h^H}{Lh^L}$ leads to within structural change because there are more high-skilled people who demand more finance subsector Z_2 services. A lower r has the same effect like a higher \bar{e}_1 . Furthermore, a lower relative return ρ (induced by an increase of the risk premium by one percentage point) or more market completeness μ (by ten percentage points) raise the skill premium and make new financial services relatively more attractive compared to services for deposits. Finally, a fall in δ to 0.335, which leads to a lower saving rate close to 18.83% as observed on average for the time span of survey years 1995-2009, induces a decline in the skill premium and leads to smaller financial sectors.

As a second exercise, we shock our calibrated model with simultaneous shocks. This means, we shock our economy by using all the shocks in the exogenous parameters together (i.e., new average values of $\bar{H}, \bar{L}, h^H, h^L, \alpha_x, \alpha_{z_1}, \alpha_{z_2}, A_x, PT_{65}, PT^{65}, LERatio$ and r^f for time span of survey years 1995-2009). Further, we assume uniform technological progress. This means, the productivities in the Z -sectors develop identical to the productivity in the X -sector. Discount parameter δ and completeness measure μ are held fixed at the calibrated values. With this procedure, we get a quantitative model prediction which can then be compared with the empirical development (see Table 1.4): Under simultaneous

shocks our model predicts a rise in the skill premium ω from 1.63 to 1.86 and two-fold structural change towards and within finance with a rise of Ψ from 5.08% to 5.21% and a rise of Φ from 13.99% to 15.02%.

Table 1.4: Predictions

Variables	Model	Data	Source	Description
ω^*	1.86	1.85	CPS	Skill premium
Ψ	5.21%	5.54%	CPS	Between sectoral structure
Φ	15.02%	23.41%	CPS	Within sectoral structure

Notes: ω^* is the equilibrium skill premium (per hour worked). Ψ corresponds to $\frac{p_{z_1}D + p_{z_2}F}{X}$ in the model and to $\Psi_{normalized}$ in the data. Φ corresponds to $\frac{p_{z_1}D}{p_{z_2}F}$ in the model and to $\Phi_{normalized}$ in the data.

Comparing the model values with data, we see that the simulated equilibrium values underestimate the between structural change (only a little) and mainly the within structural change. This means, additional shocks are needed to come closer to data values. According to our analysis, possible candidates for such additional shocks (unobserved in our data) are: More market completeness (μ -shock) and biased technical change in the Z_2 -sector (shown in Table 1.3) or diminished fixed costs in the financial sector and distorted portfolio choices as discussed in Section 1.7.

Overall, the simulated development in our calibrated model illustrates the channels that lead to the observed rise in the skill premium and the two-fold structural change towards and within the financial sector fairly well; at least as far as these changes are caused by economic fundamentals. As pointed out in the beginning of this section, the normalized financial sector ratios considered here are amplified in reality by rents.

1.10 Conclusion

The presented 3x3-model of goods production and financial services helps to explain the two-fold structural change towards and within the financial sector. The analysis emphasized demand side effects by using quasi-homothetic preferences of the Stone-Geary form and accounted for supply side effects by considering for different skill-intensities in production of goods and financial services. The theoretical analysis was based on established building blocks for modeling a multi-sector economy with production and was at the same time sufficiently tractable to allow analytical results. The comparative-static equilibrium analysis showed the effects of productivity progress and technical change, skill supply, present and future subsistence requirements as well as financial product innovation on the skill premium and on the sectoral structure of an economy. Both the

size of the financial sector relative to the non-financial sector as well as the size of the new finance sector relative to the traditional finance sector were considered. Moreover, in several extensions the robustness of the results was discussed and the effects of rents or distortions in the financial sector were addressed. The main insight of the results from the theoretical analysis can be summarized as follows: If one looks for a single economic source (apart from assuming rents or distortions) that could explain the twofold structural change towards and within finance and the rising skill premium simultaneously, the income effect is a robust candidate. Other channels, like relative price effects within the financial sector lead to more ambiguous results.

The quantitative results derived in the theoretical analysis were illustrated quantitatively by calibrating the model to U.S. data from the CPS of 1980-1994. The numerical implementation of the model shows that the subsequent development observed in the CPS for 1995-2009 can be explained fairly well. While uniform productivity growth, working through the income effect, is confirmed as a main source of structural change towards and within finance, skill biased technical change is important too for matching the rise in the skill premium.

The paper leaves open two main questions which are important in the current debate about real economic development and financialization. The first open problem is the finance premium. While it is obvious that the rents revealed by the premium contribute to inequality and blow up the structural change towards and within finance considered in this paper, the question where the premium comes from is less clear. In recent years, several attempts have been made to explain the premium by asymmetric information between shareholders and employees in the banking sector. Yet, this can only explain the redistribution of earnings within the financial sector. Our hypothesis is that it is the asymmetry between financial agents and their clients which allows extracting rents. After all, the financial sector is an expert system to start with. It would be worthwhile to integrate this aspect into the presented framework; possible channels for modeling the rent-generating information asymmetry would be intransparent cost structures or confusion by financial innovation (distorted μ -beliefs).

The second open question left to future research is how structural change towards and within the financial sector affects economic productivity. The literature on financial development and growth has identified market completion by financial innovation as an important source of growth. Does the recent evidence on a negative effect of financial development on economic growth indicate that the huge flood of new financial products since the 1990s has not really completed markets but rather generated obfuscation? In the framework presented in this paper such confusion would induce euphoric beliefs about the degree of market completeness μ which is one of the drivers of structural change within

finance and at the same time a possible lever for rent extraction. Another possible channel for a growth dampening effect could be the absorption of high-skilled labor in the finance sector, which leads to scarcity of talent outside the financial sector and may slowdown productivity growth.

To take stocks of this paper: The empirical evidence shows that the expansion of the financial sector and the changing structure within the financial sector towards new finance are partly caused by the finance premium. This is a rent which remains unexplained in the presented paper. But there are also economic fundamentals which drive the twofold structural change. These drivers are the focus of the paper. The main explanation for the observed two fold structural change is a rise in average income generated by uniform productivity growth across sectors and factors, which changes demand for financial services, combined with skill-biased technical change that drives up the skill premium.

Could the structural change towards and within finance, accompanied by a rise in the skill premium, come to a halt? According to our model, apart from a slowdown of growth, the following factors exert downward pressure on finance shares and skill premium: Finance-biased productivity progress, less attractive risky investments, a decline in the saving rate or a stop in the proliferation of new financial products.

2 An equilibrium model with diversification-seeking households, competing banks and (non-)correlated financial innovations

2.1 Introduction

The financial sector has experienced innovation dynamics with an immense increase in the number of financial products over the last decades (see Figure 2.1). This proliferation in the number of financial products raises two main questions: First, what are potential determinants of the number of financial products? Second, what are the consequences of the increase in the number of financial products for the aggregate output and the size of the banking sector? This paper contributes to an explanation of the number of financial products in an equilibrium model with diversification-seeking households and competing banks and it looks at possible consequences of new financial products.

To demonstrate the empirical relevance of the analysis of the paper, Figure 2.1 provides some motivating evidence. It shows the expansion of the number of financial products traded on different exchanges over the last decades. The graphs give a general impression of the rise in the number of products in the financial sector. A closer look shows that the number of financial products listed at the Swiss Exchange SIX increased tenfold in the last two decades (from 3,190 financial products in 1995 to 34,888 in 2014). For the Deutsche Börse we observed a rise from 25,133 different financial products in 2002 to 1,416,712 in 2014. The number of financial products traded at the London Stock Exchange increased by 26% in eight years (from 16,292 in 2006 to 20,571 in 2014). At the Shanghai Stock Exchange SSE the number of listed securities rose from 30 in 1990 to 2,786 in 2013. Although, the increase in the number of financial products varies among the exchanges in time and absolute numbers, the trend seen at these exchanges is observed globally. It can be taken as a strong indicator for the financial innovation dynamics, even though data on exchange traded financial products only partly mirror the whole development (e.g., over-the-counter traded financial products are neglected). In this paper the number of financial products (not the specific characteristics of financial innovations nor the dynamics of the development or cross country differences) is the key variable considered.

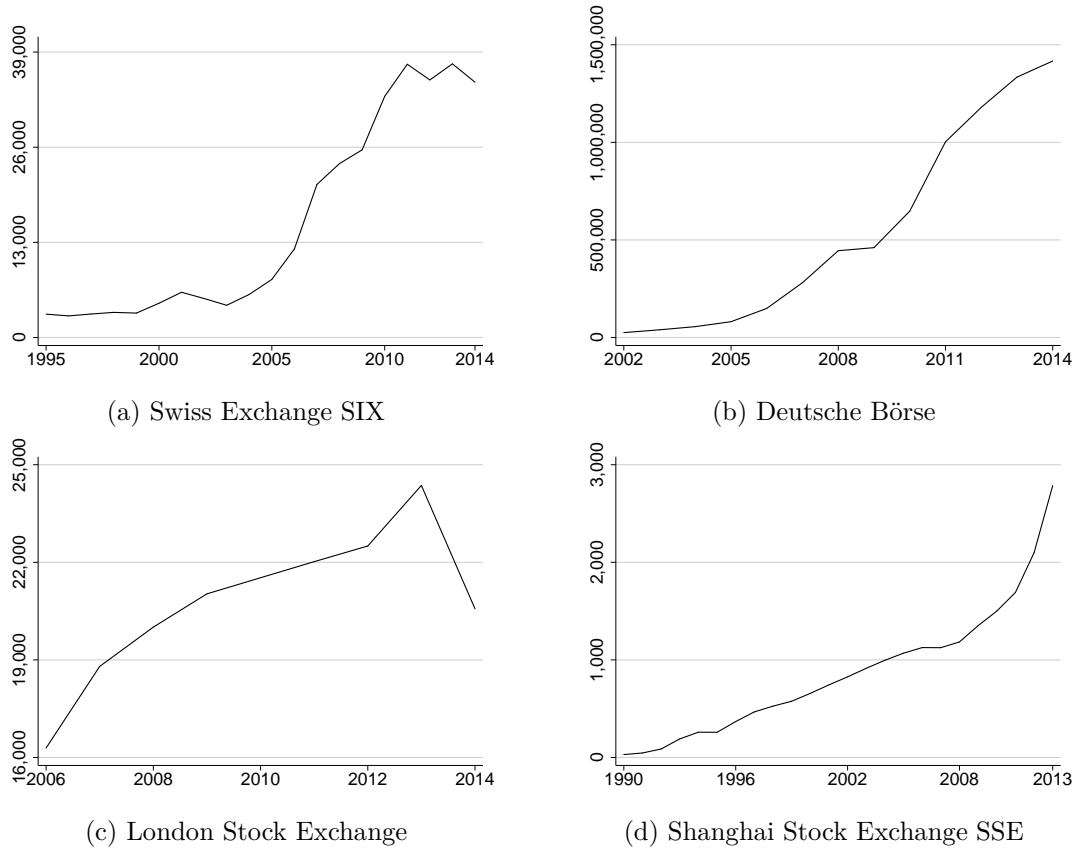


Figure 2.1: Number of financial products traded on exchanges

Notes: The graphs show the number of financial products (i.e., shares, bonds and securitized derivatives) listed at (a) the Swiss Exchange SIX, (b) the Deutsche Börse, (c) the London Stock Exchange and (d) the Shanghai Stock Exchange SSE, respectively, over the last decades. Sources: SIX, FESE, ECB and SSE (see bibliography for details).

The main contribution of this paper is to provide a theoretical general equilibrium framework, which identifies fundamentals and wrong beliefs as potential drivers of new financial products. The proposed model has the following key modeling elements: First, there is demand for and supply of the number of financial products and a fee at which they are provided. The demand comes from households (“financial investors”) whose main motive is diversification. The supply is provided by financial sector firms (“banks”) which are characterized by a cost structure and acting in a competitive market. Like in any market, the equilibrium is determined by preferences, cost parameters and competition. Second, financial products are tied to the production sector, where the returns of the financial products are generated. Third, risky financial products are divided into two categories of assets: On one hand, there is a set of independent financial products with real projects as underlyings (“financial innovation” in the proper sense of market completion) and on the other hand, there are correlated derivatives derived from the first set. Thereby, possible drivers for the growth of the second category are neglected correlation on the demand side or cheaper costs of provision on the supply side. Within this framework the modeling

strategy focuses on the most essential mechanisms: (i) a diversification motive on the demand side captured by a mean-variance approach; (ii) a convex cost function of financial product supply with fixed and variable cost components which capture the two aspects of financial services (i.e., the provision of financial products and the management of financial portfolios); and (iii) free entry competition (zero-profit condition) in the banking sector.

This model structure allows determining the equilibrium number of financial products and the equilibrium fee charged by banks explicitly. The model's key findings of the comparative-static effects of fundamentals on the number of financial products are intuitive and in line with empirical observations in a qualitative sense: From the demand side, new financial products are stimulated if more households act as financial investors, if they are less risk averse or if a larger volume of wealth is invested. From the supply side, financial products are fostered if the quality of risky investment projects serving as real underlyings rises (higher return or lower risk), banks costs decline or their efficient size increases.

In addition to the effects of the fundamentals on the number of financial products, this paper analyzes the consequences of correlation between financial products in connection with erroneous household beliefs about this correlation. It is found that neglected correlation can boost new financial products so that correlated derivatives lead to a waste of resources if they imply costs without adding diversification possibilities. Additionally, correlated derivatives bias saving decisions if households neglect part of the correlation between them, which decreases households' welfare. Furthermore, it is elaborated how banks endogenously adapt their supply behavior of financial products if households neglect correlation. The model shows that if households have wrong beliefs about the correlation of financial products, banks are "cheating" by providing correlated derivatives instead of financial products based on real projects. New financial products are thus not always the result of changes in the fundamentals. They can also be induced by correlation neglecting investors and a cost-minimizing "cheating" behavior of banks. These results allow illustrating an immense increase in the total number of financial products. At the normative level, these results support policy recommendations (e.g., patenting or rating of financial innovations) which aim to protect households, the real economy and the banking sector from the negative effects of correlated financial products.

Finally, the model is used to evaluate consequences of more financial products on aggregate output and the financial sector at the macro-level. Specifically, the theoretical analysis can explain a positive relationship between the aggregate output level, its volatility, the size of the banking sector and the number of financial products.

This research project is related to several fields of economic research. It connects insights from macroeconomics, portfolio choice, security design and microeconomics of

banking with financial innovation.

The object of the presented paper is to endogenize the number of financial products.¹ This is related to the literature on financial innovation and security design, which emphasizes that new financial products foster diversification possibilities and risk sharing (see, for example, Allen and Gale (1988, 1991, 1994), Duffie and Jackson (1989), Duffie and Rahi (1995), Kero (2013), Miller (1986), Pesendorfer (1995), Ross (1989), Simsek (2013b) or Tufano (2003)). To my knowledge, none of these previous studies explicitly considers simultaneously fundamentals from the real economy and the financial sector as well as wrong households' beliefs as determinants of the number of financial products.

I analyze the topic from a static macroeconomic perspective by basing financial products on real investment projects. The returns of financial products are generated by real investment projects with technological uncertainty in the sense of Diamond (1967). Similar to Acemoglu and Zilibotti (1997) a financial innovation occurs whenever a new risky financial product based on a novel, independent real investment project is offered. In addition, the model will be expanded by introducing correlated derivatives which are derived from other financial products rather than being based on real investment projects as underlyings. The results of C  lerier and Vall  e (2014), who find that complexity of structured products and their interlinkage increased over the last decade, motivate the introduction of such correlated financial products.

I take a mean-variance approach for both the return structure of financial products and household preferences (as initiated by Markowitz (1952)). This is a reasonable alternative to expected utility maximization in state-space, as used in Acemoglu and Zilibotti (1997) and others, because it allows focusing on the first two moments of the return distribution.²

This paper adds to standard macroeconomic literature by embedding an explicit financial intermediation sector into a model with real investment projects and households who save. Thereby, I assume the Cone (1983) / Jacklin (1986)-result which states that producers and households cannot act in the financial sector directly but they need a bank which is intermediating and matching their borrowing wants and investments needs.³ The general approach to connect households and real investment projects with financial intermediation (under potentially distorted beliefs) is affine to Falkinger (2014). However, he works with a state-space uncertainty structure with focus on imperfect knowledge and he does not

¹For an empirical discussion of financial innovations see Frame and White (2004), Miller (1986) and Tufano (2003), who give an overview over the development of types of financial innovations.

²Mean-variance optimization coincides with expected utility maximization for special cases (shown among others by Levy and Markowitz (1979); Merton (1969); Samuelson (1969) or Tobin (1985)): If utility is quadratic and returns are normally distributed or under CRRA utility and log-normally distributed returns.

³A historical overview on financial intermediation is provided by Allen and Santomero (1998). The literature is surveyed by Gorton and Winton (2003).

endogenize the number of financial products. Additionally, compared to Falkinger (2014), this paper models the technology structure and competition in the banking sector explicitly. Thereby, compared to other authors who impose costs for financial intermediation and innovation, I assume a more detailed cost function of banks:⁴ The presented model incorporates variable portfolio management costs proportional to the complexity of the financial portfolio and two types of fixed “innovation” costs; namely convex handling costs and linear issuing costs which allow differing between financial products based on real underlyings and correlated derivatives. Using this cost structure, the equilibrium number of financial products and the fee is determined by perfect bank competition which leads to zero-profits. Although banks are characterized by a cost structure and are assumed to act in a competitive market – rather than seen as a black box – many issues from the finance literature (e.g., oligopoly structure or rents in the banking sector, liquidity provision or bubbles) are neglected.

This paper assesses wrong beliefs as driver of new financial products. More precisely, I explore the effects of neglected correlation (whereas, for example, Simsek (2013a,b) looks at belief disagreement between financial investors or Gennaioli et al. (2012, 2013) analyze neglected tail risk). The relevance for considering erroneous assessment of correlations by investors is supported by experimental studies by Eyster and Weizsäcker (2011) and Kallir and Sonsino (2009). They find that subjects in the role of financial investors neglect correlation between assets. The results of this paper arising from neglected correlation as driver of the emergence of correlated derivatives, question the finding of Pesendorfer (1995) who shows potential welfare enhancing effects of redundant financial products.

Finally, this paper deals with the aggregate consequences of financial innovations.⁵ The findings of a co-movement of more financial products and aggregate output are in line with the financial development argumentation in Acemoglu and Zilibotti (1997). For them the fact that new financial assets can be used as a tool for diversifying the risk of high-productive technologies (“safety in variety” (Acemoglu and Zilibotti, 1997, p.718)) is an important source of growth. In contrast to Acemoglu and Zilibotti, who find that aggregate variability can be decreasing with more securities being available, in the model presented here portfolio and aggregate output volatility increase with the number of financial products because the effect of more capital being invested riskily dominates the diversification effect. This is similar to the “hedge-more/bet-more-effect” in Simsek (2013b,

⁴For example, Freixas and Rochet (2008, p. 70) assume managing costs per volumes of deposits or loans; Allen and Gale (1988, 1991) assume a (linear) fixed cost for issuing new financial claims; Bisin (1998) works with variable and fixed cost depending on the payoff vector of securities; or Pesendorfer (1995) uses marketing costs proportional to traded volumes of financial products and per customer.

⁵For a more normative perspective on financial innovations see Haliassos (2013). He collects views from academia and practice on the question whether financial innovations can be blamed for the recent economic crisis and concludes that there was not too much but too little financial innovation.

p.1367) and the “volatility paradox”-result of Brunnermeier and Sannikov (2014, p.379) saying that securities that foster risk sharing can lead to more volatility. Further, the model can predict a co-movement of the number of financial products and the growing size of the financial sector (i.e., financialization) as observed in data.⁶

The rest of the paper is structured as follows: In Section 2.2, the theoretical model is described. The general equilibrium and comparative-static effects of changes in fundamentals are analyzed in Section 2.3. Section 2.4 expands the baseline model by incorporating correlated financial products and erroneous household beliefs. In particular, the effects arising from such distortions on the equilibrium number of financial products and household utility are discussed. Section 2.5 provides insights into banks’ supply behavior of both correlated derivatives and financial products based on real underlyings. Section 2.6 provides model intuition how the empirical observed exponential increase in the total number of financial products can arise. Section 2.7 discusses consequences of financial innovations for households, aggregate output and the size of the banking sector. Section 2.8 concludes.

2.2 Model

The model consists of three parts: First, real investment projects with return generating technologies which are the basis for financial products; second, households as financial investors who can choose to invest in a safe asset and in risky financial products; third, banks which offer the financial products and manage financial portfolios.

2.2.1 Financial products

2.2.1.1 Projects with mean-variance return generating technology

Financial products are based on real investment projects which generate returns. There are two types of investment projects with linear technologies in capital: (i) a safe and (ii) a potential set $\{j|j \in [1, \infty)\}$ of risky return generating technologies. The safe investment project s has output $Y_s = rK_s$ and yields on capital input K_s a deterministic return $\partial Y_s / \partial K_s = r$ with $\text{var}(r) = \sigma_r^2 = 0$. The output of the risky investment project j is $Y_j = R_j K_j$. Its return $\partial Y_j / \partial K_j = R_j$ is stochastic and describes the idiosyncratic productivity of capital input K_j . This means that the riskiness comes from uncertainty in the return generating technology (i.e., technological uncertainty in the sense of Diamond (1967)). Risky investment projects are assumed to be independent. Risky return gen-

⁶See Epstein (2005), Falkinger et al. (2015), Greenwood and Scharfstein (2013) or Philippon and Reshef (2012) for a discussion of financialization.

erating technologies are i.i.d. and described by their mean and variance: The expected return of a risky investment project, $\mathbb{E}(R_j) = \mathbb{E}(R)$, and the corresponding variance, $\text{var}(R_j) = \sigma^2$, is the same for all j . Define $\tilde{R} \equiv \mathbb{E}(R) - r$ as the excess return of the risky projects. In line with the conventional return-risk trade-off the following assumption holds:

Assumption 2.1. $\tilde{R} > 0$ and $\sigma^2 > 0$.

Aggregate output Y in this economy corresponds to the sum of output of the safe and the N open risky investment projects. (It is explained in Section 2.2.3.3 how the number N is determined in equilibrium.) Expected aggregate output is given by

$$\mathbb{E}(Y) = rK_s + \mathbb{E}(R) \sum_{j=1}^N K_j \quad (2.1)$$

and its volatility captured by the variance is

$$\sigma_Y^2 = \text{var}(Y) = \text{var} \left(rK_s + \sum_{j=1}^N R_j K_j \right) = \sigma^2 \sum_{j=1}^N K_j^2. \quad (2.2)$$

2.2.1.2 Design of financial products and diversification effect

Real investment projects receive capital by serving as underlying of financial products. Hence, financial products reflect the property rights on the distribution of returns generated by the investment projects. Project returns are given back as payoffs of financial products to households, who act as financial investors. A safe asset is based on the safe return generating technology and promises the deterministic return r . Risky financial products are based on risky real investment projects.⁷ That means, a financial innovation is designed in a stylized way and defined as follows: A financial innovation occurs whenever a bank decides to offer a new risky financial product by accessing and opening up for their clients another independent investment project.⁸ It promises a stochastic return with expectation $\mathbb{E}(R)$ and variance σ^2 .

By construction, N risky financial products correspond to N open independent, risky projects in which investments are made. The N financial products generate independent streams of returns. Therefore, each financial innovation raises the level of diversification. The N risky financial products can be combined into a bundle B . The expected return

⁷Sections 2.4 and 2.5 models the consequences of relaxing this assumption.

⁸This process is similar to Acemoglu and Zilibotti (1997). However, in their case it is not the banks who offer financial products for a fee which covers the costs of providing financial products and managing financial portfolios. They have agents who run production projects with a minimum size requirement and compete for funds by issuing (cost-free) financial products.

of a bundle is the weighted sum of each individual asset's expected return and its variance depends on the weights, the variances of the individual assets and their respective covariances. With uncorrelated assets, there is the following diversification effect:

Lemma 2.1. *In a bundle B of N financial products maximal possible diversification is achieved if equal amounts are invested in each of the N independent, risky financial products. The resulting bundle B promises an expected return $\mathbb{E}(R_B) = \mathbb{E}(R)$ and variance $\sigma_B^2 = \frac{\sigma^2}{N} < \sigma^2$.*

Proof. Equal amounts means a share $\frac{1}{N}$ per risky asset. It follows $\mathbb{E}(R_B) = \sum_N \frac{1}{N} \mathbb{E}(R) = N \frac{1}{N} \mathbb{E}(R) = \mathbb{E}(R)$. And $\sigma_B^2 = \sum_N \frac{1}{N^2} \sigma^2 = N \frac{1}{N^2} \sigma^2 = \frac{\sigma^2}{N}$. Since the variance is scaled by the square of the weights, the variance of a weighted sum is minimized if the weights are $\frac{1}{N}$. \square

Lemma 2.1 means, the more N there are, the lower becomes the variance σ_B^2 of the return of the bundle B with still the same expected return $\mathbb{E}(R_B)$. Hence, N represents the “level of diversification” in the risky bundle B . If the number of financial innovations $N \rightarrow \infty$ all idiosyncratic risk would diversified away and markets became complete.

2.2.2 Households' mean-variance portfolio choice

The economy consists of I households, who act as financial investors. Households have endowments $w > 0$ and save it for later consumption by investing into a financial portfolio (PF) and paying a fee τ to banks. They decide how to invest w to maximize utility from later consumption of the returns on their investments. In line with the portfolio choice problem in the finance literature, households do this by mean-variance optimization (originated by Markowitz (1952)). They are assumed to maximize

$$\mathbb{E}(R_{PF}) - \frac{\gamma \sigma_{PF}^2}{2} - \tau, \quad (2.3)$$

where $\mathbb{E}(R_{PF})$ is the expectation and σ_{PF}^2 is the variance of households' financial portfolio unit return. $\gamma > 0$ reflects their risk aversion. $\tau > 0$ denotes the unit fee charged by the bank. This fee is deducted from the return on households' portfolio (similar to the “intermediation margin” in Freixas and Rochet (2008, p.73)). Households have to pay the fee per unit of endowment to the banks for the provision of financial products and the managing of their portfolios.⁹ τ is endogenously determined by the banking sector in equilibrium and will depend on the complexity of the financial portfolio (see Section 2.2.3). It is assumed that households cannot observe banks' pricing function.

⁹Assume that the parameters are such that even in the worst case realization of the financial portfolio unit return, households have enough endowment to cover the fee.

The mean-variance approach captures the key characteristics of a concave utility function. It makes evident that risk averse agents wish to gain a high expected return, to have a low-risk portfolio and to pay a small fee. By counterbalancing return and risk, households invest in the safe asset and in the bundle B of the N risky financial products. More specifically, given a level of diversification N , households choose the share α of their endowment w to be invested in the risky bundle B . The rest, $(1 - \alpha)$ of w , is invested in the safe asset. The expected unit return on their financial portfolio is thus $\mathbb{E}(R_{PF}) = \alpha\mathbb{E}(R) + (1 - \alpha)r$. The corresponding variance is $\sigma_{PF}^2 = \alpha^2\sigma_B^2 + (1 - \alpha)^2 0 = \alpha^2\sigma_B^2$. Substituting these values in (2.3), we have the following decision problem of households:

$$\max_{\alpha \in [0,1]} \alpha\mathbb{E}(R) + (1 - \alpha)r - \frac{\gamma\alpha^2\sigma_B^2}{2} - \tau, \quad (2.4)$$

where $\alpha \in [0, 1]$ excludes short-selling and reflects the budget constraint. The maximization problem in (2.4) gives the optimal share α^* invested in the risky bundle B :¹⁰

$$\alpha^*(N) = \frac{\mathbb{E}(R) - r}{\gamma\sigma_B^2} = N \frac{\tilde{R}}{\gamma\sigma^2}. \quad (2.5)$$

The second equality uses $\sigma_B^2 = \frac{\sigma^2}{N}$ as derived in Lemma 2.1. Equation (2.5) shows that the optimal share $\alpha^*(N)$ depends linearly on the level of diversification N . The unit return of the optimal financial portfolio generated by this portfolio choice has mean $\mathbb{E}(R_{PF}^*) = N \frac{\tilde{R}}{\gamma\sigma^2} \mathbb{E}(R) + (1 - N \frac{\tilde{R}}{\gamma\sigma^2})r = N \frac{\tilde{R}^2}{\gamma\sigma^2} + r$ and variance $\sigma_{PF}^{2*} = (N \frac{\tilde{R}}{\gamma\sigma^2})^2 \sigma^2 = N \frac{\tilde{R}^2}{\gamma^2\sigma^2}$. Both the expectation and the volatility of the portfolio unit return depend positively on N . A more pronounced diversification effect in B (i.e., more N) leads to more risky high-return investments and thus to higher expected return and variance of the portfolio unit return.

Substituting the optimal share $\alpha^*(N)$ from (2.5) into (2.4), I get the indirect utility function

$$v^*(N, \tau) = \mathbb{E}(R_{PF}^*) - \frac{\gamma\sigma_{PF}^{2*}}{2} - \tau = N \frac{\tilde{R}^2}{2\gamma\sigma^2} + r - \tau. \quad (2.6)$$

The households' utility $v^*(N, \tau)$ depends on N , the fundamentals and the fee τ . It is linearly increasing in N . Since more of financial products N promises higher utility, households are denoted to be diversification-seeking. This mean that, other things equal, households are attracted by financial innovation.

¹⁰In the following I focus on interior solutions. For $\gamma\sigma^2 < \infty$, $\alpha^* = 0$ will never be optimal because of $\mathbb{E}(R) > r$. The upper bound $\alpha^* = 1$ does not change, but only dampen the model predictions because with a fix $\alpha^* = 1$ more N do lower the variance (due to the diversification effect from Lemma 2.1), but do not increase the expectation of the portfolio unit return. Furthermore, household finance data (see e.g., Carroll (2002)) show that $\alpha^* < 1$ is far more realistic than $\alpha^* = 1$.

Lemma 2.2 summarizes the two intermediary results.

Lemma 2.2. *Given a level of diversification N , household's portfolio choice is $\alpha^*(N)$ with α^* increasing in N . Households are diversification-seeking because the indirect utility $v^*(N, \tau)$ is increasing in N .*

Proof. Follows directly from equations (2.5) and (2.6). □

2.2.3 Banking sector

Banks act as financial intermediaries and financial innovators. Their task is to transform savings into productive investment capital by offering N financial products and managing financial portfolios. Banks financially innovate when issuing more financial products by accessing and opening up for their clients more real investment projects. As described above, banks offer such financial innovations to households who seek diversification.

2.2.3.1 Technology structure

Banks are symmetric and operate with the same technology structure, which is characterized by the efficient bank size and the cost parameters.

An exogenous efficient bank size $I^* > 0$ is assumed, which means that I^* households can be served per bank. Under the assumption that each household is only served by one bank, it follows that the number of households I divided by the efficient bank size I^* determines the number of banks $n = \frac{I}{I^*}$ in the economy.¹¹

Banks face costs for providing financial products and intermediation services. The key assumption on banks is:

Assumption 2.2. *Banks are symmetric and a bank's total costs $C_{tot}(N)$ are increasing and convex in N .*

This is in line with the intuition that providing more financial products and managing more diverse portfolios costs more. Consider the following specification as a concrete cost structure fulfilling Assumption 2.2:¹² Total costs $C_{tot}(N)$ consist of variable costs

¹¹Symmetry means that banks are not differentiated and all offer the same variety of financial products $[1, N]$ by assumption. It could also be understood as a representative bank. The symmetry assumption is justified by the empiric: It is actually observed that all banks are offering more or less the same financial products (e.g., made accessible by an initial public offering) and that households are likely to be served by only one “house bank”. One could determine I^* endogenously by assuming an oligopoly with free entry in the banking sector. In this case, I^* would be fixed by the minimum value of the natural oligopoly properties of a customer relationship cost function $F(I/n)$.

¹²Note that by Assumption 2.2, $C_{tot}(N)$ fulfills the usual assumption of convexity and regularity of banks' costs function (see, for example, Freixas and Rochet (2008, p.70) for general motivation of convex

of managing the volume of savings and of fixed innovation costs of handling and advertising financial products. Variable portfolio management costs are proportional to the complexity of the financial portfolio: Variable costs of managing risky invested savings are $c^r > 0$, whereas safe invested savings impose variable costs $c^s > 0$ with $c^s < c^r$. c^r and c^s arise to the bank per volume of households' total saving in the risky bundle, $\alpha^*(N)wI^*$, and the safe asset, $(1 - \alpha^*(N))wI^*$, respectively. Thus, total variable costs are $c^r\alpha^*(N)wI^* + c^s(1 - \alpha^*(N))wI^*$. Define $\tilde{c} \equiv c^r - c^s > 0$ as the excess costs of risky investments. For positive asset demand the cost of managing an asset must be smaller than the returns it yields (i.e., $c^r < \mathbb{E}(R)$, $c^s < r$ and $\tilde{c} < \tilde{R}$). In addition to the variable costs, there are two types of fixed costs for offering risky financial products (innovation costs). The first can be interpreted as handling costs of a variety N of independent financial products. They occur to a bank when accessing and opening up for their clients N risky real investment project. These handling costs are convex increasing in N . This reflects the fact that with more N , it costs banks disproportionately more to be careful in providing new, independent financial products. Without loss of generality these costs are assumed to be dN^2 , where $d > 0$ is a constant. Second, for each risky financial product there are fixed costs $f > 0$. These can be interpreted as advertising costs or as the issuing fee of a financial product. Total advertising costs are fN . Bank's total costs are therefore

$$C_{tot}(N) = c^r\alpha^*(N)wI^* + c^s(1 - \alpha^*(N))wI^* + dN^2 + fN. \quad (2.7)$$

$C_{tot}(N)$ is in line with Assumption 2.2 because $\alpha^*(N)$ in (2.5) and thus the total variable costs are linear increasing in N and the fixed costs, $dN^2 + fN$, are convex increasing in N .

2.2.3.2 Profit function

Banks charge to each client a fee τ per unit of managed wealth w to compensate for the portfolio management and the supply of financial products. The total fee a household has to pay is thus proportional to the volume of savings w that the bank has to manage. Total fee income per bank is then $I^*w\tau$. By deducting total costs from the total revenue raised by the fee, the banks' profit function becomes

$$\Pi(N, \tau, \alpha^*(N)) = I^*w\tau - C_{tot}(N), \quad (2.8)$$

costs; however, not in relation to financial products N). $C_{tot}(N)$ shall be understood in broad sense; it includes entry costs, managers' compensation, reserves and so on. This cost structure allows the discussion of comparative-static effects of cost parameters on the equilibrium (see Section 2.3.2) and the introduction of correlated derivatives (see Sections 2.4 and 2.5).

where $C_{tot}(N)$ defined by (2.7) uses $\alpha^*(N)$. It shows that a bank's profit depends on the fee τ it charges, the number of financial products N it offers, households' behavior $\alpha^*(N)$ and technology parameters.

2.2.3.3 Banking sector equilibrium: Competition and optimal offer

Banks compete by offering the safe asset and N risky financial products for a fee τ . It is assumed that there is perfect competition by free entry (i.e., competition for the market). Free entry prevents any bank from gaining positive profits in a banking sector equilibrium. If this did not hold, banks would be substituted away by new market entrants. In line with the contestable market concept banks will enter and overtrump others (by offering more N or ask a lower fee τ) as long as they can gain positive profits. More specifically, banks are offering to households a financial service package characterized by (N, τ) , which must fulfill two conditions in equilibrium: First, for not being overtrumped the package must not yield positive profit to the bank. Second, as a participation constraint for households the package offered by a bank must generate for households at least the utility level they would get from any other zero-profit package.¹³ Lemma 2.3 presents formally the two conditions which the equilibrium-package (N^*, τ^*) must fulfill.

Lemma 2.3. *Let $\Pi(N, \tau, \alpha^*(N))$ be the bank's profit function from equation (2.8) and let $v^*(N, \tau)$ be the household's indirect utility from equation (2.6). In equilibrium, banks offer a (N, τ) -package such that*

$$\begin{aligned} \Pi(N, \tau, \alpha^*(N)) &= 0 \\ v^*(N, \tau) &> v^*(N', \tau') \quad \forall \quad N', \tau'. \end{aligned} \tag{2.9}$$

Proof. If banks offer (N, τ) -packages not fulfilling $\Pi(N, \tau, \alpha^*(N)) = 0$, there is market entry by another bank which offers $N' > N$ or $\tau' < \tau$ taking over the clients and still gaining $\Pi(N', \tau', \alpha^*(N')) > 0$. This continues until $\Pi(N, \tau, \alpha^*(N)) = 0$. If banks do not offer utility optimizing (N, τ) -packages to households, they are outplayed by a market entrant who offers a (N', τ') -package with $v^*(N', \tau') > v^*(N, \tau)$ and takes over the clients (if $\Pi(N', \tau', \alpha^*(N')) > 0$ finally being displaced by another bank with $\Pi(N'', \tau'', \alpha^*(N'')) = 0$). Thus, at the end $\Pi(N, \tau, \alpha^*(N)) = 0$ and $v^*(N, \tau) > v^*(N', \tau')$. \square

To determine the number of financial products N^* and the corresponding τ^* simultaneously as a package in equilibrium, banks' supply decision is combined with households'

¹³Note that this competition structure yields the same equilibrium outcome as a situation in which an auctioneer would do a profit maximization for the n banks by exploiting households' utility surplus fully.

demand decision (portfolio choice): From the profit function (2.8) follows that in a zero-profit banking sector equilibrium the fee τ charged for providing a specific number of financial products N is given by

$$\hat{\tau}(N) \equiv \frac{C_{tot}(N)}{I^*w}, \quad (2.10)$$

where $C_{tot}(N)$ uses $\alpha^*(N)$. For each potential level of diversification N , banks determine with $\hat{\tau}(N)$ the necessary fee to be charged to have zero profits. Since $C_{tot}(N)$ is increasing and convex in N by Assumption 2.2, it holds that $\hat{\tau}'(N) > 0$ and $\hat{\tau}''(N) > 0$. Thus, a more complex bundle B consisting of more financial products N makes banks charge a higher fee to households.

In addition, the equilibrium condition (2.9) states that the equilibrium N^* and τ^* must fulfill $v^*(N^*, \tau^*) > v^*(N', \tau')$ for any N' and τ' fulfilling (2.10). The equilibrium number of financial products N^* is implicitly determined by the following equation

$$\frac{\partial v^*(N, \hat{\tau}(N))}{\partial N} = 0, \quad (2.11)$$

where $v^*(N, \hat{\tau}(N))$ is the indirect utility function evaluated at the equilibrium fee $\hat{\tau}(N)$ for N . It states that, in equilibrium, for households the diversification utility of more N must correspond to marginal fee costs of more N .

Proposition 2.1. *Suppose the assumptions on households' utility and on banks' technology structure and free entry competition hold with parameters satisfying the inequality $2f\gamma\sigma^2 < I^*\tilde{R}(\tilde{R} - 2\tilde{c})w$. Then, there exists a unique equilibrium (N^*, τ^*) -package. It is characterized by:¹⁴*

$$N^* = \frac{I^*\tilde{R}(\tilde{R} - 2\tilde{c})w - 2f\gamma\sigma^2}{4d\gamma\sigma^2} \quad (2.12)$$

$$\tau^* = \frac{16c^s d + \frac{I^*(\tilde{R}^2 - 4\tilde{c}^2)\tilde{R}^2 w}{\gamma^2 \sigma^4} - \frac{8\tilde{c}f\tilde{R}}{\gamma\sigma^2} - \frac{4f^2}{I^*w}}{16d} \quad (2.13)$$

Proof. See Appendix B.1.1 for derivation. □

Figure 2.2 gives an illustration of the equilibrium conditions (2.10) and (2.11) and the determination of the equilibrium in the (N, τ) -space. It shows the zero-profit locus which is given by the $\hat{\tau}(N)$ -function defined in equation (2.10). $\hat{\tau}(N)$ is increasing and

¹⁴For $N^* > 0$ the assumption that $f\gamma\sigma^2 < \frac{I^*\tilde{R}(\tilde{R}-2\tilde{c})w}{2}$ is necessary. Furthermore, $N^* \geq 1$ is guaranteed if $\frac{I^*w\tilde{R}^2}{2\gamma\sigma^2} \geq 2d + f + \frac{\tilde{R}I^*w\tilde{c}}{\gamma\sigma^2}$.

convex in N . The household's utility from different (N, τ) -packages is represented by the indifference curve in the (N, τ) -space. From equation (2.6) follows that it is a straight line with slope $\frac{\bar{R}^2}{2\gamma\sigma^2} > 0$. According to condition (2.11), the equilibrium (N^*, τ^*) -package is determined by the unique point where the household's indifference curve is tangent to the $\hat{\tau}(N)$ -function. This point is consistent with both banks' zero-profit condition and households' optimal portfolio choice. In contrast any other (N, τ) -package does not jointly fulfill the two equilibrium conditions given in (2.9).

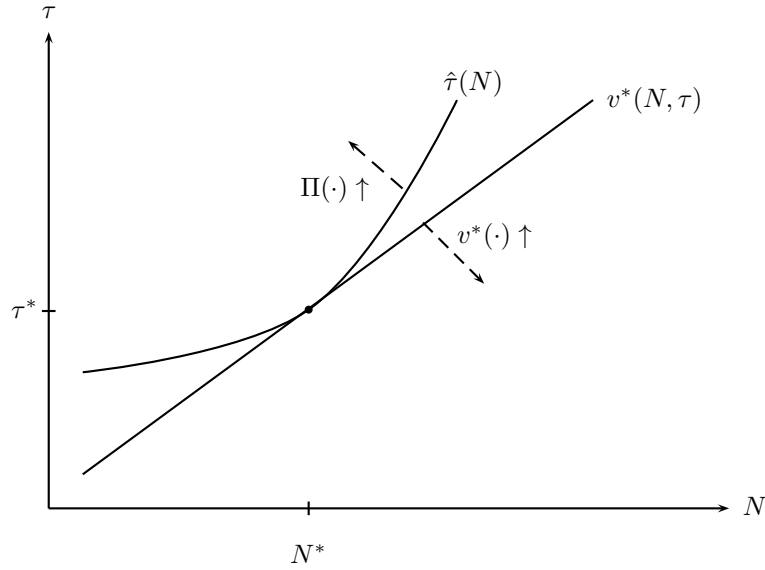


Figure 2.2: Equilibrium package with N^* and τ^*

Notes: The figure plots a bank's zero-profit fee function and a household's indifference curve both for $\alpha^*(N)$. The tangency point fulfills the equilibrium conditions and determines the equilibrium number N^* of financial products and the corresponding fee τ^* .

2.3 Equilibrium analysis

2.3.1 Equilibrium characterization

The equilibrium in this economy is defined as follows:

Definition 2.1. *An equilibrium is a number of financial products and a fee for financial services $\{N, \tau\}$; a household's portfolio choice $\{\alpha\}$ and the resulting capital allocations $\{K_s, K_j\}$, $j = 1, \dots, N$, which is consistent with household optimization and a zero-profit equilibrium in the banking sector.*

The equilibrium in this economy is characterized by N^* and τ^* defined in (2.12) and (2.13), respectively, by

$$\alpha^*(N^*) = N^* \frac{\tilde{R}}{\gamma \sigma^2}, \quad (2.14)$$

$$K_s^*(N^*) = (1 - \alpha^*(N^*))wI = (1 - N^* \frac{\tilde{R}}{\gamma \sigma^2})wI \quad (2.15)$$

and

$$K_j^*(N^*) = \frac{\alpha^*(N^*)}{N^*}wI = \frac{\tilde{R}}{\gamma \sigma^2}wI, \quad j = 1, \dots, N^*. \quad (2.16)$$

The equilibrium given by equations (2.12)-(2.16) allows now answering the questions raised in the introduction about the determinants and the consequences of financial innovations. More precisely, it enables the discussion of the equilibrium effects of changes in the fundamentals on the number of financial products (see comparative statics on N^* in Section 2.3.2). Further, it allows analyzing effects of more financial products on aggregate variables (see Section 2.7).

2.3.2 Effects of fundamentals on financial innovations

The equilibrium solution N^* presented in (2.12) depends on the fundamentals from the underlying return generating technologies \tilde{R} and σ^2 , households' endowment w and risk attitude γ and the bank's cost parameters \tilde{c}, d, f as well as the number of households I^* served per bank. The signs of the comparative-static effects of the fundamentals on the number of financial products are shown in Table 2.1 and summarized in Proposition 2.2.¹⁵

Table 2.1: Comparative statics of fundamentals

Parameter	I^*	w	γ	\tilde{R}	σ^2	\tilde{c}	f	d
on N^*	+	+	-	+	-	-	-	-

Proposition 2.2. *Under the assumptions of Proposition 2.1, in equilibrium financial innovations are stimulated by a larger volume of wealth managed per bank (higher I^* or more w), less risk averse households (lower γ), more attractive risky financial products (higher \tilde{R} or lower σ^2) or higher cost-efficiency in the banking sector (lower \tilde{c} , f or d).*

Proof. Follows from equation (2.12): For $N^* > 0$ we have $\frac{\partial N^*}{\partial I^*} > 0$, $\frac{\partial N^*}{\partial w} > 0$, $\frac{\partial N^*}{\partial \gamma} < 0$, $\frac{\partial N^*}{\partial \tilde{R}} > 0$, $\frac{\partial N^*}{\partial \sigma^2} < 0$, $\frac{\partial N^*}{\partial \tilde{c}} < 0$, $\frac{\partial N^*}{\partial f} < 0$ and $\frac{\partial N^*}{\partial d} < 0$. \square

¹⁵The comparative-static effects of the fundamentals on the equilibrium fee τ^* , portfolio share α^* and capital allocations K_s^* , K_j^* are presented in Table B.1 in Appendix B.2.

By qualitatively cross-checking the results in Proposition 2.2 with data, I find that these comparative-static results are intuitive and in line with empirical observations: First, the data shows that the safe return r , represented by the effective federal funds rate in the U.S., dropped from around 10% in the 1980s to 0.1% in 2014.¹⁶ According to the model a lower r raises, *ceteris paribus*, \tilde{R} and lead to a higher number of financial products. The reasoning behind this is that households are looking for more yield and therefore ask for risky, high-return financial products. Second, wealth volumes w managed per bank increased dramatically throughout the recent decades (for example through leveraging). According to Proposition 2.2 this drives more financial products because innovation costs are better shared. Additionally, the empiric shows that the financial sector was characterized by consolidation with a decrease in the number of banks in the U.S. from above 35,000 in 1980 to around 15,000 in 2009.¹⁷ A decreased number of banks corresponds in the presented model to a higher efficient bank size I^* . This stimulates more financial products because a higher I^* allows banks to serve a given N at a lower τ per household. This induces demand for more N . Finally, as Philippon and Reshef (2012) argue, bank deregulation (as observed in the last decades) can intensify innovation activity of banks because it lowers a bank's innovation costs by loosening the frontier for financial innovations. Furthermore, numbers of Philippon (2015) on the unit costs of financial intermediation (i.e., total income of financial intermediaries divided by the amount of intermediated assets) hint at a decrease of bank cost in the short run: His calculations show a decrease in the quality adjusted unit costs of financial intermediation from 1.8% in 1990 to 1.5% in 2012. Such a decrease can be a result of diminished variable management costs \tilde{c} or of lower handling costs d or advertising costs f (e.g., through IT-progress). Any such reduction of costs triggers, according to the model, new financial products.

Hence, the model developed here identifies reasonable channels and provides an explanation for new products in the financial sector by identifying fundamentals as possible drivers. Qualitatively these drivers are consistent with empirical trends.

2.4 Correlated financial products and neglected correlation

I now address the question whether new financial products can also be induced by (wrong) beliefs – in contrast to changes in the fundamentals. More precisely, I allow neglected correlation to be an additional channel which could facilitate new financial products. So

¹⁶Data taken from the Federal Reserve Bank of St.Louis. See bibliography for details.

¹⁷Data taken from OECD.Stat. See bibliography for details.

far, financial innovations were new financial products based on real investment projects with independent return generating technologies. In fact, however, financial innovations are partly just combinations of existing financial products and exhibit correlation among them.¹⁸ Henceforth, I call such products correlated derivatives. In the further analysis the baseline model is extended to account for such products. The question of interest is how correlated financial products affect the equilibrium number of financial products in this economy and households' welfare – the answer depends on whether households are able to see the respective correlation or not.

2.4.1 Correlated derivatives

Assume that banks do not thoroughly check whether the financial products they offer are all based on independent real investment projects. In fact, they are offering derivatives D which are correlated with the “real” financial innovations N . The design of derivatives is assumed to be done in a simplistic way: For each of the N financial products based on independent real projects banks offer $\nu \geq 0$ other products based on the same underlying. They promise the same return. Hence, in addition to N banks offer $D = \nu N$ cloned derivatives with correlation $\rho = 1$. In sum, banks supply a total $M = N + D = (1 + \nu)N$ of financial products which have, in fact, only N independent real investment projects as underlyings. For now, the multiplier ν is taken as exogenous parameter. (In Section 2.5, ν is endogenized through a cost-minimizing behavior of banks.) For the provision of a correlated derivative the following is assumed:

Assumption 2.3. *The provision of a correlated derivative requires no handling cost d , yet, it has to be issued at cost f .*

Under Assumption 2.3 a bank's innovation costs become $dN^2 + f(1 + \nu)N$ and total costs are $C_{tot}(N, \nu, \epsilon) = c^r \alpha^*(N, \nu, \epsilon) w I^* + c^s (1 - \alpha^*(N, \nu, \epsilon)) w I^* + dN^2 + f(1 + \nu)N$, where $\alpha^*(N, \nu, \epsilon)$ is determined in the next subsections. Compared to the baseline model the changes are that issuing fixed costs increase to $(1 + \nu)f$ instead of f and that households' portfolio choice $\alpha^*(N, \nu, \epsilon)$ might be affected by ν ; depending on whether households neglect correlation ϵ between N and D or not.

¹⁸C  l  rier and Vall  e (2014) analyze the development of complexity of structured products and find, in fact, increased interlinkage between them.

2.4.2 Neglected correlation

2.4.2.1 No neglect

First, for the sake of comparison, the effects of correlated derivatives on the equilibrium are analyzed for the case in which households are fully able to differ between the financial products N based on independent real investment projects and the cloned derivatives D . They are aware of the correlation $\rho = 1$ of the clones and understand that the derivatives D are not contributing to diversification. This means that households base their portfolio decision still on the true bundle variance. Thus, $\sigma_B^2(N, \nu) = \sigma_B^2 = \frac{\sigma^2}{N}$ is invariant to ν , where $\sigma_B^2(N, \nu)$ denotes the variance of a bundle B with $M = (1 + \nu)N$ assets with equal weights.¹⁹ Households' portfolio decision remains unbiased and the same fraction $\alpha^*(N, \nu, \epsilon) = \alpha^*(N)$ of wealth is allocated to risky assets as in the baseline. The only thing that changes compared to the baseline is the increase of issuing cost f to $(1 + \nu)f$ and thus the fee which is consistent with zero profits rises from $\hat{\tau}(N)$ to $\hat{\tau}(N, \nu) \equiv \hat{\tau}(N) + \frac{\nu N f}{I^* w}$.

By applying the same procedure as in Section 2.2 to determine the equilibrium, we obtain for the total number of financial products

$$M_D^* = (1 + \nu)N_D^* = (1 + \nu) \frac{I^* \tilde{R}(\tilde{R} - 2\tilde{c})w - 2(1 + \nu)f\gamma\sigma^2}{4d\gamma\sigma^2}, \quad (2.17)$$

which is a multiple of the number of financial innovations N_D^* based on real projects.²⁰ Compared to N^* in (2.12), the equilibrium number of real financial innovations N_D^* is now lower. The reason is that the issuing costs per financial product based on a real investment project rise from f to $(1 + \nu)f$ because also its clones have to be issued and advertised (“cost effect”). Thus, the multiplier ν has the same comparative-static effect as an increase in f . The correlated derivatives waste resources in the sense that they incur extra costs without adding value in the sense of diversification possibilities. Derivatives substitute real financial innovations. Due to the ν -multiplier, derivatives increase the total number of financial products compared to the frictionless model whenever $M_D^* > N^*$.²¹ In sum, if $I^* w$ is large compared to f , then derivatives stimulate the total number of financial products through this “multiplication effect”.

¹⁹Actually, households are in their portfolio choice indifferent (if they have to cover all costs anyhow) between investing a certain fraction of their wealth symmetrically in the N basic financial products or in all $M = (1 + \nu)N$ products which are offered. It is assumed without loss of generality that they do the latter.

²⁰See Appendix B.1.1 for the derivation and Table B.2 in Appendix B.3.1 for a full characterization of the other equilibrium variables.

²¹ $M_D^* > N^* \Leftrightarrow (1 + \nu)N_D^* > N^*$ reduces to the following condition: $I^* \tilde{R}(\tilde{R} - 2\tilde{c})w > 2f\gamma\sigma^2(2 + \nu)$. Thereby, $I^* \tilde{R}(\tilde{R} - 2\tilde{c})w > 2f\gamma\sigma^2(1 + \nu)$ is already needed for $N_D^* > 0$.

2.4.2.2 (Partial) correlation neglect

Second, I consider the more interesting case with erroneous beliefs; namely, neglected correlation. I analyze what happens if households are not (fully) able to differentiate between the N financial products with independent projects as underlyings and the D correlated derivatives. Assume that households neglect a part $\epsilon \in [0, 1]$ of the correlation $\rho = 1$ and base their portfolio decision on $\tilde{\rho} \equiv 1 - \epsilon$. This inaccurate belief can either arise if households have limited ability to differ between financial products or if banks are not providing them with all relevant information. The erroneous assessment of ρ makes households overestimate the diversification effect provided by a bundle B of equally weighted $M = N + D$ assets. With belief $\tilde{\rho}$, the *perceived* variance of such a bundle B is:²²

$$\sigma_B^2(N, \nu, \epsilon) = \frac{1 + \nu(1 - \epsilon)}{(1 + \nu)N} \sigma^2 \leq \frac{1}{N} \sigma^2 = \sigma_B^2 \quad (2.18)$$

For $\epsilon = 0$, we are back in the case $\sigma_B^2(N, \nu, \epsilon) = \sigma_B^2$ in which households account correctly for the fact that D are perfectly correlated clones of N . In contrast, if $\epsilon > 0$ households overestimate the diversification effect and hence underestimate the variance of the risky portfolio. Then neglected correlation leads to a biased portfolio choice $\alpha^*(N, \nu, \epsilon) = \frac{\tilde{R}}{\gamma \sigma_B^2(N, \nu, \epsilon)} = \frac{(1 + \nu)N}{1 + \nu(1 - \epsilon)} \frac{\tilde{R}}{\gamma \sigma^2}$. This says, with increased neglect the households excessively invest in the risky bundle B . Thus, in addition to the increased fixed costs from f to $(1 + \nu)f$, we have biased portfolio choices with more risky invested savings increasing also variable costs. This rises the zero profit fee $\hat{\tau}(N, \nu, \epsilon) \geq \hat{\tau}(N, \nu) > \hat{\tau}(N)$ further.

With the same procedure as in Section 2.2, the equilibrium total number of financial products $M_{D,NC}^*$ for the case with derivatives and neglected correlation is determined:²³

$$M_{D,NC}^* = (1 + \nu)N_{D,NC}^* = (1 + \nu) \frac{I^* \tilde{R} (\tilde{R} - 2\tilde{c}) w (1 + \nu) - 2(1 + \nu)(1 + \nu(1 - \epsilon)) f \gamma \sigma^2}{4d\gamma \sigma^2 (1 + \nu(1 - \epsilon))} \quad (2.19)$$

Like in the $\epsilon = 0$ case, the cost of issuing the ν -fold clones of the financial products based on real projects raises the zero-profit fee. This “cost effect” would result in a lower $N_{D,NC}^*$ compared to the baseline N^* in (2.12). The neglect of correlation distorts the equilibrium provision of financial products through an additional channel; namely, through the variance bias $\frac{1 + \nu(1 - \epsilon)}{1 + \nu}$. This bias raises the households’ demand for risky financial

²²See derivation in Appendix B.1.2.

²³See Appendix B.1.1 for the derivation and Table B.2 in Appendix B.3.1 for a full characterization of the other equilibrium variables.

products and counteracts the cost effect of ν on $N_{D,NC}^*$. We have that the higher the correlation neglect ϵ , the higher is the number of real financial innovations $N_{D,NC}^*$. This is because households believe to have more diversification and invest a larger share of their savings in risky financial products (“neglected correlation effect”). $N_{D,NC}^* > N^*$ from (2.12) as soon as the neglect ϵ is above a threshold $\bar{\epsilon}$,²⁴ because then the “neglected correlation effect” outweighs the “cost effect” of ν on $N_{D,NC}^*$. In addition, derivatives increase the total number of financial products because of the ν -fold multiple of each financial product based on a real investment project. For $M_{D,NC}^* > N^*$ we have that the creation of derivatives in connection with neglected correlation increases the total number of financial products.²⁵ Indeed, the number of financial products $M_{D,NC}^*$ can rise dramatically by an interaction of the “neglected correlation effect” and the “multiplication effect”. This shows that the proliferation of the number of financial products (as illustrated in Figure 2.1) can be driven by correlated financial products in connection with erroneous diversification assessment and need not always to be the result of changes in the fundamentals.

2.4.3 Consequences of correlated derivatives and neglected correlation

Correlated derivatives in connection with correlation neglect are not only affecting the total number of financial products in equilibrium, but have further consequences for households: In sum, there are two effects of correlated derivatives in connection with wrong beliefs about correlations, which affect the individuals’ welfare. First, the issuing costs of cloned financial products must be covered by higher fees. Second, households are misled in their portfolio choice. This means that the true utility level – generated by the biased portfolio choice under the true bundle variance $\sigma_B^2 = \frac{1}{N_{D,NC}^*} \sigma^2$ which is larger than the planned one $\sigma_B^2(N_{D,NC}^*, \nu, \epsilon)$ – is smaller than the households’ planned utility level. In other words, the overestimated diversification effect is disillusioned when confronted with reality – with an adverse effect on households’ welfare. Thus, the cost and the bias distortion both have negative effects on households’ utility.²⁶

²⁴ $\bar{\epsilon} = \frac{2f\gamma\sigma^2(1+\nu)}{2f\gamma\sigma^2\nu + I^*\tilde{R}(\tilde{R}-2\tilde{c})w}$.

²⁵ $\epsilon > \bar{\epsilon}$ is a sufficient condition for $M_{D,NC}^* > N^*$, not a necessary one. For a sharp condition follows from (2.19) and (2.12): $M_{D,NC}^* > N^*$ if and only if $I^*\tilde{R}(\tilde{R}-2\tilde{c})w(1+\nu+\epsilon) > 2f\gamma\sigma^2(1+\nu(1-\epsilon))(2+\nu)$.

²⁶In Appendix B.3.2 the equations for the equilibrium utility level from the baseline model, $v^*(N^*, \tau^*)$, and the model with derivatives but no correlation neglect, $v^*(N_D^*, \nu, \tau_D^*)$, as well as planned $v^*(N_{D,NC}^*, \nu, \epsilon, \tau_{D,NC}^*)^p$ and true $v^*(N_{D,NC}^*, \nu, \epsilon, \tau_{D,NC}^*)^t$ utility levels from the model with derivatives and neglected correlation are provided.

The consequences of correlated derivatives and neglected correlation on households' utility and the equilibrium number of financial products are summarized in Proposition 2.3.²⁷

Proposition 2.3. *If there are correlated derivatives D and households neglected part ϵ of the correlation between the derivatives and the financial products with independent projects as underlyings, portfolio choices become biased (“excessive” risky investments) and effective household utility declines. Moreover, the total number of financial products $M = N + D$ provided in equilibrium can fundamentally exceed the number issued in an undistorted equilibrium.*

Proof. Implied by equations (2.18), (2.19) and the utility levels in Appendix B.3.2. \square

These results support policy recommendations which aim to prevent the supply of redundant financial products and protect financial investors: ν should be kept small (e.g., through patenting of financial innovations) and ϵ should be minimized (e.g., through transparent information requirements, costumer services or responsible rating agencies which reveal information about financial products) so that the equilibrium without distortions can be realized.

2.5 Endogenous supply of correlated derivatives

I propose now a way to endogenize the derivative multiplier ν . This reveals a potential explanation for why banks might offer redundant financial products. The modeling procedure to determine the endogenized ν follows the one from the baseline model.²⁸ Banks are competing for diversification-seeking households by providing the optimal level of diversification at the lowest possible fee costs. Households are seeking for diversification possibilities in order to invest in risky financial products with high returns without increasing their risk exposure. According to (2.5), the optimal share households invest in the bundle B of risky assets is $\alpha^* = \frac{\tilde{R}}{\gamma\sigma_B^2}$. Without diversification (i.e., if there is just one risky asset) we have $\sigma_B^2 = \sigma^2$. With diversification the bundle variance is lowered to $\sigma_B^2 = \frac{1}{a}\sigma^2$, where $a \geq 1$ defines the degree of diversification offered by bundle B . The portfolio allocation is thus $\alpha^*(a) = a\frac{\tilde{R}}{\gamma\sigma^2}$. In the baseline analysis the level of diversification $a = \frac{\sigma^2}{\sigma_B^2}$ was given by N . In contrast, under neglected correlation the *perceived* level of

²⁷In addition to the consequences of correlated derivatives and neglected correlation on households' utility and the equilibrium number of financial product, aggregate output and the size of the banking sector are affected. See Section 2.7 for a discussion of the effects of financial innovations on aggregates.

²⁸For notational simplicity, I suppress D, NC as the sub-index of M and N for the remainder of the paper, but I keep on discussing the model version with correlated derivatives and neglected correlation (if not stated otherwise).

diversification is given by $a = \frac{(1+\nu)N}{1+\nu(1-\epsilon)}$ (see (2.18)), which is a function of the correlation neglect ϵ .

Specifically, the procedure for endogenizing ν is as follows: First, banks determine the minimal fee for each potential level of perceived diversification a by combining in a cost-minimizing way financial innovations based on real underlyings N and correlated derivatives $D = \nu N$ in the bundle B (see Section 2.5.1). This allows them to deliver a at the lowest possible fee. Any higher fee level would be underbid by a potential competitor. Second, in general equilibrium, this supply behavior of banks is combined with households' mean-variance preferences so that the equilibrium level of perceived diversification a^* which promises households the highest possible utility level can be determined (see Section 2.5.2).

2.5.1 Cost-minimal provision of perceived diversification

For each level of perceived diversification a banks offer the combination of N financial products based on real underlyings and the derivative multiplier ν leading to $D = \nu N$ derivatives which minimizes total costs $C_{tot}(N, \nu, \epsilon)$. Thereby, they take into account that households base their portfolio choice $\alpha^*(a)$ on the perceived level of diversification a which is influenced by the correlation neglect ϵ . For any a with $\alpha^*(a) \in [0, 1]$ the bank's cost-minimizing combination of N and ν is determined by the following optimization problem:

$$\begin{aligned} \min_{\nu, N} \quad & c^r \alpha^*(a) w I^* + c^s (1 - \alpha^*(a)) w I^* + d N^2 + f(1 + \nu) N \\ \text{subject to} \quad & a = \frac{(1+\nu)N}{1+\nu(1-\epsilon)} \\ & \alpha^*(a) = a \frac{\bar{R}}{\gamma \sigma^2} \\ & N \geq 1, \nu \geq 0 \end{aligned} \tag{2.20}$$

The first constraint of the optimization problem defines a locus, which describes all (N, ν) -combinations providing the same perceived level of diversification a (i.e., iso-diversification curve). It can be rewritten:²⁹

$$N(\nu, a, \epsilon) = \frac{1 + \nu(1 - \epsilon)}{1 + \nu} a \tag{2.21}$$

By substituting this into (2.20), I determine implicitly the cost-minimizing ν^* for a given

²⁹See Figure B.1 in Appendix B.1.3.1 for the iso-diversification curve. Note that for $\epsilon > 0$ $\frac{\partial N(\nu, a, \epsilon)}{\partial \nu} < 0$ and $\frac{\partial^2 N(\nu, a, \epsilon)}{\partial \nu^2} > 0$, $\frac{\partial N(\nu, a, \epsilon)}{\partial a} > 0$ and $\frac{\partial N(\nu, a, \epsilon)}{\partial \epsilon} < 0$.

perceived level of diversification a .³⁰

$$\frac{f}{2da} = \epsilon \frac{\frac{1}{1-\epsilon} + \nu^*}{(1 + \nu^*)^3} \quad (2.22)$$

The analytical examination of the program in (2.20) and its solution in (2.22) provide insights into banks' supply behavior of financial products based on real underlyings and of correlated derivatives:

Equation (2.22) determines the multiplier ν^* only for $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$ (i.e., inner solutions if the constraints $\nu^* \geq 0$ and $N^* \geq 1 \Leftrightarrow \nu^* \leq \frac{a-1}{1-a(1-\epsilon)}$ by (2.21) are non-binding) and has a unique positive solution for this range of ϵ .³¹ For the whole range of $\epsilon \in [0, 1]$ banks set the multiplier ν^* as follows: First, it is at its minimum $\nu_{\min}^* = 0$ for $0 \leq \epsilon \leq \underline{\epsilon}$; second, it is given by (2.22) for $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$; and third, it is determined by $\nu^* = \frac{a-1}{1-a(1-\epsilon)}$ for $1 \geq \epsilon \geq \tilde{\epsilon}$ (with maximum $\nu_{\max}^* = \frac{a-1}{1-a(1-\tilde{\epsilon})}$ at $\tilde{\epsilon}$ and with $\nu^* = a-1$ for $\epsilon = 1$). This means in words: If households are capable to distinguish between the financial products based on real underlyings and the derivatives and to assess their correlation well enough with $0 \leq \epsilon \leq \underline{\epsilon}$, banks do not clone the financial products based on real underlyings – although households might not consider the correlation fully – because the provision of correlated derivatives is costly and they do not add much to the perceived diversification. Thus, $\underline{\epsilon} \geq 0$ is the lower bound of neglected correlation beyond which banks start “cheating” on households by providing correlated derivatives. For high correlation neglect with $1 \geq \epsilon \geq \tilde{\epsilon}$ banks are providing the full perceived level of diversification a with derivatives based on only one real financial innovation. Hence, $\tilde{\epsilon} \leq 1$ is the upper bound of neglected correlation after which banks are providing the perceived level of diversification a with only one financial product based on a real investment project, but many clones of it. These boundaries for ν^* imply by (2.21) that the number of financial products based on independent real investment projects, N^* , starts at maximum $N_{\max}^* = a$ for $0 \leq \epsilon \leq \underline{\epsilon}$ and ends in its minimum $N_{\min}^* = 1$ for $1 \geq \epsilon \geq \tilde{\epsilon}$.

If households neglect part $\tilde{\epsilon} > \epsilon > \underline{\epsilon}$ of the correlation, banks duplicate the financial products based on independent real investment projects and create correlated derivatives because the latter's “innovation” is cheaper. For these positive inner solutions follow comparative-static effects of changes in $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$, a , d and f on the derivative multiplier (given by (2.22)) and on the number of financial products based on real underlyings (given

³⁰See Appendix B.1.3.1 for the derivation.

³¹For the argumentation on uniqueness of a positive ν^* for $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$ see Appendix B.1.3.1. $\underline{\epsilon}$ and $\tilde{\epsilon}$ are derived from equation (2.22) by setting $\nu^* = 0$ and $\nu^* = \frac{a-1}{1-a(1-\epsilon)}$, respectively. $\underline{\epsilon} = \frac{f}{2da+f}$ decreases in a . It starts at $\frac{f}{2d+f}$ for $a = 1$ and is limited at $\lim_{a \rightarrow \infty} \underline{\epsilon} = 0$. $\tilde{\epsilon} = \frac{4(a-1)d+af+\sqrt{f}\sqrt{8(a-1)d+a^2f}}{2a(2d+f)}$ increases in a . It starts at $\frac{f}{2d+f}$ for $a = 1$ and is limited at $\lim_{a \rightarrow \infty} \tilde{\epsilon} = 1$. See Appendix B.1.3.2 for derivations.

by $N^*(\nu^*, a, \epsilon)$ in (2.21)). (2.23) summarizes the comparative statics where + is a positive and - a negative comparative-static effect.³²

$$\nu^*(\underset{+}{\epsilon}, \underset{+}{a}, \underset{+}{d}, \underset{-}{f}) \quad \text{and} \quad N^*(\underset{-}{\epsilon}, \underset{+}{a}, \underset{-}{d}, \underset{+}{f}) \quad (2.23)$$

As indicated by (2.23), for $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$ the multiplier ν^* rises with the level of correlation neglect ϵ . Hence, the more correlation the households neglect, the more banks can deceive them by selling redundant financial products. (2.23) indicates further that for $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$ the number of financial innovations based on real underlyings, N^* , decreases with the level of correlation neglect ϵ . The reason of the decrease of N^* in ϵ is twofold: In addition to a direct negative effect of the correlation neglect ϵ , there is an indirect effect through the increased multiplier ν^* that makes fewer financial products based on independent real investment projects necessary to achieve a . These comparative-static effects mean, if households neglect a larger part ϵ of the correlation between financial products banks are providing the level of perceived diversification a by substituting financial products based on real underlyings through cloned derivatives (ϵ -effect).

Further, (2.23) shows that for $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$ an increase in the perceived level of diversification a raises both ν^* and N^* because it shifts the iso-diversification curve outwards. This means, for a higher perceived level of diversification more financial products based on real investment projects are offered along with the rising number of cloned derivatives (a -effect).

Finally, (2.23) summarizes for $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$ that ν^* rises with the handling costs d of real investment products because with a higher d the supply of derivatives becomes relatively cheaper. N^* decreases in d due to the inverse argumentation. ν^* declines (and N^* rises) as the issuing cost f increases because this makes derivatives relatively more costly to offer.

For the comparative statics of the exogenous on the total number of financial products $M^* = (1 + \nu^*)N^*$ follows for $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$:³³

$$M^*(\underset{+/-}{\epsilon}, \underset{+}{a}, \underset{+}{d}, \underset{-}{f}) \quad (2.24)$$

The effect of the neglect ϵ on M^* depends on the elasticity of the multiplier ν^* with respect to ϵ : $\frac{\partial M^*}{\partial \epsilon} \geq 0 \Leftrightarrow \frac{\partial \nu^*}{\partial \epsilon} \frac{\epsilon}{\nu^*} \geq \frac{\epsilon}{1-\epsilon}$. The condition for a positive effect of ϵ on M^* is fulfilled for low ϵ , but not necessarily if the correlation neglect is high. Thus, the total of financial products M^* is hump-shaped in ϵ : Since ν^* is increasing and N^* is decreasing in ϵ , their product $(1 + \nu^*)N^*$ is larger for middle values than it is closer to the boundaries. Further,

³²See Appendix B.1.3.3 for derivations.

³³See Appendix B.1.3.4 for derivations.

the total number of financial products is clearly increasing in the level of diversification a because both ν^* and N^* are increasing in a . Finally, M^* increases in d and decreases in f as ν^* does.

Banks' partial equilibrium supply behavior of financial products derived from (2.20) and summarized in (2.23) and (2.24) provides insights why banks might offer redundant (or correlated) financial products: If financial investors make an erroneous diversification assessment the creation of correlated financial products permits banks to stay competitive because it allows them to satisfy investor's perceived diversification demand a at lower costs.

2.5.2 Equilibrium with neglected correlation

To determine the general equilibrium, the supply behavior of banks is combined with households' mean-variance preferences. For each potential level of perceived diversification a , banks have determined the cost-minimizing derivative multiplier ν_a^* and the number of financial products N_a^* based on real underlyings. They fixed the corresponding zero-profit fee $\hat{\tau}_a(N_a^*, \nu_a^*)$ which just covers their total costs. Then, in equilibrium, for a given level of neglect ϵ we get the equilibrium level of perceived diversification $a^* = \frac{(1+\nu_a^*)N_a^*}{1+\nu_a^*(1-\epsilon)}$, which promises households the highest utility level given their optimal portfolio choice $\alpha^* = a^* \frac{\bar{R}}{\gamma \sigma^2}$ and the corresponding equilibrium fee $\tau_{a^*}^*(N_{a^*}^*, \nu_{a^*}^*)$.

I obtain the general equilibrium by numerical simulation.³⁴ The main results can be summarized as follows: The higher the correlation neglect ϵ , the higher is the equilibrium level of perceived diversification a^* . A higher ϵ shifts banks' cost function for a specific level of perceived diversification a down because the issuing of D is cheaper than the provision of N .³⁵ Since households willingness to pay for a is not affected by ϵ , this leads to a demand for a higher a^* . Under more neglect households are willing to hold many financial products because they believe that each of them provides diversification possibilities. Thus, in general equilibrium the multiplier ν^* increases in ϵ twofold: First, because banks can "cheat" more by satisfying demanded perceived diversification with cheap correlated clones (ϵ -effect) and second because more diversification is asked (a^* -effect). In addition, as long as ϵ is not too high, N^* is increasing in ϵ because more diversification is demanded (i.e., the positive a^* -effect outweighs the negative ϵ -effect on N^*). In sum, the total number

³⁴The simulation procedure is described in Appendix B.4.1. Figure B.2 in Appendix B.4.2 gives simulated results of the equilibrium in dependence of the correlation neglect $\epsilon \in [0, 1]$. In panels (a)-(f) a^* , ν^* , N^* and M^* and the true σ_B^2 and planned bundle variance $\sigma_B^{2*}(N^*, \nu^*, \epsilon)$ as well as the fee τ^* are plotted. Note that the corner solutions of ν^* as discussed above are hardly visible in the simulation because $\underline{\epsilon}$ is close to 0 and $\bar{\epsilon}$ approaches 1 for increased levels of a^* .

³⁵See Appendix B.1.3.5 for the derivations of comparative-static effects on the cost function.

of financial products M^* rises with the level of correlation neglect ϵ .³⁶ The consequence of these results is that the underestimation of the true bundle variance σ_B^{2*} is increasing in ϵ because there is (i) more neglect ϵ of (ii) more perceived diversification a^* . Hence, planned and true bundle variance are diverging. Further, the demand for more a^* implies a higher fee τ^* as long as ν^* and N^* are increasing in ϵ . If the negative ϵ -effect outweighs the positive a^* -effect on N^* it is possible that the equilibrium fee τ^* decreases with ϵ because banks do not need to provide many costly financial innovations based on real underlyings anymore.

These general equilibrium results show that the observed product innovations in the financial sector can be driven by correlated financial products; namely, if banks compete for diversification-seeking investors who (partially) neglect correlation. New financial products can be induced by investors' neglected correlation and the subsequent caused cost-minimizing "cheating" behavior of banks and must not always be the result of changes in the fundamentals.

2.6 Intuition for immense increase in the number of financial products

With these results from the general equilibrium analysis it can be illustrated how a series of the static model (with increasing correlation neglect) is able to intuitively replicate the immense increase in the number of financial products as seen in Figure 2.1.

2.6.1 Empirical features

When taking a deeper look at, for example, the data from the Swiss Exchange SIX shown in Figure 2.1 one observes some interesting features which are presented in Figure 2.3. The figure shows the number of financial products listed at the Swiss Exchange SIX from 1995-2014 subdivided into security types bonds, shares and structured products. In this time range the number of bonds varied between 2,019 and 1,235 and the number of shares decreased from 530 in 1995 to 289 in 2014. However, the most remarkable change can be observed in the number of structured products: It increased from 551 in 1995 to 32,896 in 2014. Thus, while in 1995 two-thirds of the total of 3,190 financial products were bonds, in 2014 94% of the total of 34,888 were structured products. This indicates that the

³⁶Panel (c) in Figure B.2 shows that beyond some ϵ -threshold the ϵ -effect dominates and N^* declines. Still the total number M^* of financial products rises (panel (d)) because the multiplier ν^* is enhanced by the rising ϵ (panel (b)).

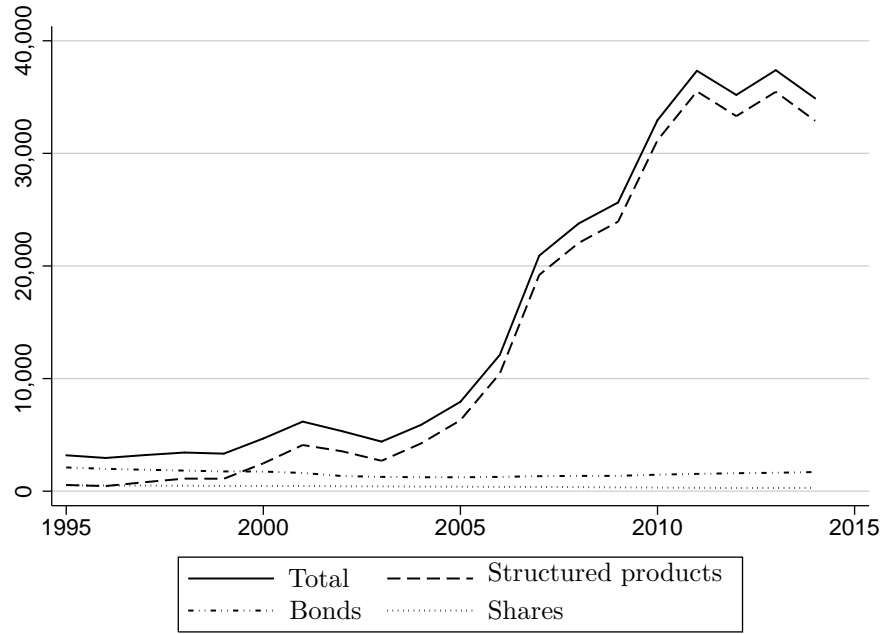


Figure 2.3: Number of financial products traded at SIX

Notes: Number of financial products listed at the Swiss Exchange SIX from 1995-2014. Shares include *Swiss shares* and *foreign shares*. Bonds include *domestic CHF bonds* and *foreign CHF bonds*. Structured products include *investment funds*, *sponsored funds*, *ETFs*, *ETSFs*, *ETPs* and *structured products and warrants*. During 2002 to 2004, there were security type reallocations (e.g., *structured products bonds* were reallocated from *bonds* to *derivatives*) and before 2003, the numbers included equities listed on the main market only. Source: SIX (see bibliography for details).

strong innovation dynamics in the financial sector arise from a dramatic increase in the number of structured products and not from basic financial products.³⁷

2.6.2 Model intuition

The presented static model can provide intuition for the empirical features of the immense increase in the number of financial products as seen in Figure 2.1 and 2.3. For that, consider several single realizations $i \in \{0, 1, \dots, 6\}$ of the presented model with varying correlation neglect ϵ_i (i.e., comparative statics with respect to ϵ). The single model realizations are ordered such that ϵ_i is incrementally increased (i.e., $0 = \epsilon_0 < \epsilon_1 < \dots < \epsilon_6$). Figure 2.4 provides an illustration of this series of single static model predictions. It shows an immense increase in the total number of financial products.

Note that, this specifically ordered series of single static model predictions can be interpreted as time series if one assumes that over the last decades households made more and more erroneous correlation assessments. One could claim that such an increasing

³⁷This conclusion is in line with findings of Célérier and Vallée (2014) who empirically show that financial products became more complex during the last decade.

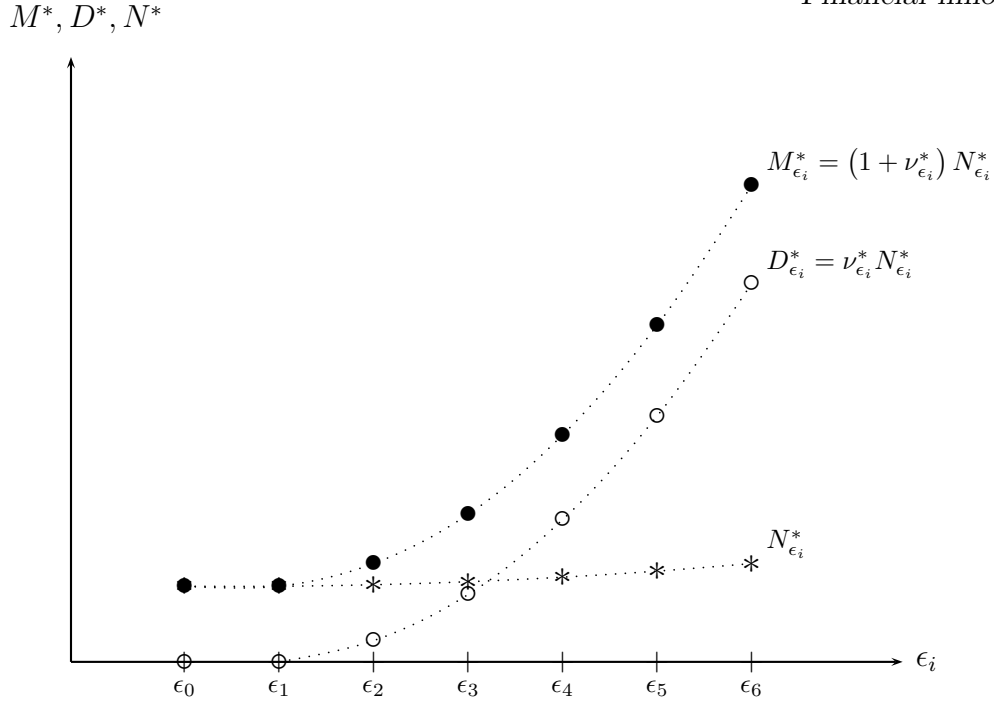


Figure 2.4: Immense increase in the number of financial products

Notes: The figure plots a series of single static model predictions for seven different levels of correlation neglect $\epsilon \in [0, 1]$ with $0 = \epsilon_0 < \dots < \epsilon_6$. For all $\epsilon_i, i \in \{0, 1, \dots, 6\}$, \bullet indicate the total numbers of financial products $M_{\epsilon_i}^* = N_{\epsilon_i}^* + D_{\epsilon_i}^* = (1 + \nu_{\epsilon_i}^*) N_{\epsilon_i}^*$, \circ indicate the numbers of correlated derivatives $D_{\epsilon_i}^* = \nu_{\epsilon_i}^* N_{\epsilon_i}^*$ and $*$ indicate the numbers of financial products based on real investment projects $N_{\epsilon_i}^*$. It shows a strong incremental increase of the total number of financial products – summing the derivatives $D_{\epsilon_i}^*$ and the real financial innovations $N_{\epsilon_i}^*$ – implied by households' increasing correlation neglect and banks' duplication behavior.

correlation neglect is caused by increased financial complexity over time if fundamentals have led to an increase in the number of financial products.

The intuition of the series of model predictions in Figure 2.4 is as follows (for the intuition, the results from the equilibrium analysis from Section 2.5.2 are used): In the model realization $i = 0$ with $\epsilon = \epsilon_0 = 0$ households do not neglect correlations so that the derivative multiplier $\nu_{\epsilon_0}^* = 0$. Thus, the model realization $i = 0$ predicts the total number of financial products to be equal to the number of financial innovations based on independent real investment projects, $M_{\epsilon_0}^* = N_{\epsilon_0}^*$. In model realizations $i \geq 1$ with $\epsilon = \epsilon_i > 0$, households make an erroneous diversification assessment and neglect more and more correlation with increasing i ($\epsilon_1 < \dots < \epsilon_6$). This increases the derivative multiplier $\nu_{\epsilon_i}^*$ incrementally due to positive ϵ - and a^* -effects on ν^* (see Section 2.5.2). Furthermore, the number $N_{\epsilon_i}^*$ of financial products based on real projects increases with ϵ_i under the assumption that the positive a^* -effect outweighs the negative ϵ -effect on N^* (see Section 2.5.2). The elevated $\nu_{\epsilon_i}^*$ and $N_{\epsilon_i}^*$ with ϵ_i result in a convex increase in the number of correlated derivatives $D_{\epsilon_i}^* = \nu_{\epsilon_i}^* N_{\epsilon_i}^*$. This means that (i) the total number $M_{\epsilon_i}^* = N_{\epsilon_i}^* + D_{\epsilon_i}^* = (1 + \nu_{\epsilon_i}^*) N_{\epsilon_i}^*$ of financial products increases immensely and that (ii) correlated derivatives drive this immense increase. These two results are in line with

empirical observations. Hence, the series of single static model predictions with increasing correlation neglect provides a model intuition for the empirical features of the exponential innovation dynamics in the financial sector.

2.7 Consequences of financial innovations

The increase in the number of financial products – inclusive the provision of correlated clones on top of financial products based on independent real investment projects – have far reaching consequences for households and macroeconomic aggregates.³⁸

2.7.1 Mean-variance effects: Diversification vs. volatility

With more financial products, $M^* = (1 + \nu^*)N^*$, a larger share of wealth is invested in risky, high-return financial products (i.e., $\alpha^* = \frac{1+\nu^*}{1+\nu^*(1-\epsilon)}N^*\frac{\bar{R}}{\gamma\sigma^2}$ increases in M^*). Thus, the equilibrium unit return on household portfolios $\mathbb{E}(R_{PF}) = \frac{1+\nu^*}{1+\nu^*(1-\epsilon)}N^*\frac{\bar{R}^2}{\gamma^2\sigma^2} + r$ increases with M^* . *Pari passu*, its true variance rises to $\sigma_{PF}^2 = (\frac{1+\nu^*}{1+\nu^*(1-\epsilon)})^2 N^*\frac{\bar{R}^2}{\gamma^2\sigma^2}$, whereas $(\frac{1+\nu^*}{1+\nu^*(1-\epsilon)})N^*\frac{\bar{R}^2}{\gamma^2\sigma^2}$ was planned. Although there is a diversification effect in the bundle B if more financial products are provided, the volatility of the portfolio rises because a more risky composition of the portfolio is chosen. This effect arises even in the baseline with $\nu^* = 0$ and $\epsilon = 0$. It is similar to the “hedge-more/bet-more-effect” in Simsek (2013b, p.1367) and Simsek (2013a). With correlated derivatives and neglected correlation $\epsilon > 0$, the overestimation of the diversification effect by households who face a mix of real financial innovations N^* and correlated derivatives $D^* = \nu^*N^*$ induces the households to take even more risk. Compared to $\epsilon = 0$, this raises the mean of portfolio unit return as expected, but the risk in the return is accompanied by more volatility than expected (note that $(\frac{1+\nu^*}{1+\nu^*(1-\epsilon)})^2 > \frac{1+\nu^*}{1+\nu^*(1-\epsilon)}$ if $\epsilon > 0$). Hence, planned and true portfolio variance diverge more if ϵ rises. This means that with correlated derivatives and neglected correlation the consequences of new financial products for the households are deceived expectations about the performance of their financial portfolios.³⁹

2.7.2 Aggregate output effects

The model is able to qualitatively replicate macroeconomic outcomes, which can be observed parallel to the innovation dynamics in the financial sector.

³⁸Note that here the impact of an increase in the number of financial products is discussed *ceteris paribus*. This means, without considering the other effects of fundamental or wrong beliefs which led to the actual increase in the number of financial products.

³⁹Similar effects are also found in Falkinger (2014) who considers erroneous probability and productivity assessments.

Furthermore, implications for the real aggregate output level and its volatility can be deduced. For the macroeconomic implications of the equilibrium with correlated derivatives and erroneous diversification assessment it is useful to look at the share α^* allocated to the risky assets. Aggregate output is determined by the factor allocations, which depend through α^* on the total number of financial products $M^* = (1 + \nu^*)N^*$. Aggregate capital allocations in the safe and the risky investment projects are $K_s^* = (1 - \alpha^*)wI$ and $K_j^* = \alpha^*wI/N^*$ for $j = 1, \dots, N^*$ (see (2.15) and (2.16)), respectively, where with derivatives and neglected correlation we have $\alpha^* = \frac{1+\nu^*}{1+\nu^*(1-\epsilon)}N^*\frac{\tilde{R}}{\gamma\sigma^2}$. For the expected aggregate output given in (2.1) I get then

$$\mathbb{E}(Y^*) = \left(\frac{1 + \nu^*}{1 + \nu^*(1 - \epsilon)} N^* \frac{\tilde{R}^2}{\gamma\sigma^2} + r \right) Iw \quad (2.25)$$

and the true aggregate volatility given in (2.2) is

$$\sigma_Y^{2*} = \left(\frac{1 + \nu^*}{1 + \nu^*(1 - \epsilon)} \right)^2 N^* \frac{I^2 w^2 \tilde{R}^2}{\gamma^2 \sigma^2}. \quad (2.26)$$

Since with more financial products a larger share of wealth is invested in high-return, risky investment projects, $\mathbb{E}(Y^*)$ and σ_Y^{2*} increase with the total number of financial products M^* . This is, more risky financial products lead to a positive aggregate output effect because the average return $\mathbb{E}(R)$ on the α^* -fraction of savings, which is increasing in M^* , is larger than the safe return r . This argumentation is in line with the literature on financial development (e.g., Acemoglu and Zilibotti (1997)). At the same time, a “volatility paradox” (Brunnermeier and Sannikov, 2014, p.379) exists:⁴⁰ More financial products that diversify risk increase aggregate volatility because they lead to more investments in the risky bundle.

Note that the rise in the output level is of linear order in the factor $(\frac{1+\nu^*}{1+\nu^*(1-\epsilon)})$, while the volatility of aggregate output is increased by the quadratic factor $(\frac{1+\nu^*}{1+\nu^*(1-\epsilon)})^2$. If the growth of financial products supplied by the financial system is (partly) driven by banks’ provision of correlated derivatives ν^* and households’ neglected correlation ϵ , then the co-movements of financial innovations and macroeconomic trends are accelerated; with the volatility of aggregate output rising more strongly than its expected level. This shows that the discussed distortions not only impact the number of financial products based on real underlying and of correlated derivatives, but also the real economy outcome.

In sum, equations (2.25) and (2.26) show, *ceteris paribus*, co-movements of expected aggregate output and aggregate volatility, with the total number of financial product.

⁴⁰Brunnermeier and Sannikov (2014) describe the “volatility paradox” as a phenomenon in which lower fundamental risk can lead to more volatility because it implies higher leverage.

These are stylized facts, which can be observed in the empirics: The aggregate output effect mirrors macroeconomic growth and the variance effect is in line with the recent rise of macroeconomic volatility shown by Carvalho and Gabaix (2013).

2.7.3 Banking sector size

The number of financial products impacts also the size of the banking sector. The size of the banking sector λ denotes the fraction of resources absorbed by banks relative to the aggregate output. The absorbed resources are measured by total fees. The size of the banking sector λ is the sum of total fees $Iw\tau$ divided by expected aggregate output $\mathbb{E}(Y)$:

$$\lambda \equiv \frac{Iw\tau}{\mathbb{E}(Y)} \quad (2.27)$$

With correlated derivatives and neglected correlation $\tau^* = \frac{C_{tot}(N^*, \nu^*, \epsilon)}{I^*w}$ and thus the relative size of the banking sector is thus given by

$$\lambda^* = \frac{\alpha^* \tilde{c} + c^s + \frac{dN^{*2} + f(1+\nu^*)N^*}{wI^*}}{\alpha^* \tilde{R} + r}, \quad (2.28)$$

where $\alpha^* = \frac{1+\nu^*}{1+\nu^*(1-\epsilon)} N^* \frac{\tilde{R}}{\gamma\sigma^2}$. The size of the banking sector depends on the total number of financial products M^* and also on the mix of products based on independent real projects N^* and correlated derivatives $D^* = \nu^* N^*$. More financial products M^* increase the share α^* of wealth allocated to risky assets. Both the resources absorbed by the variable costs (i.e., $\alpha^* \tilde{c} + c^s$ in (2.28)) and the expected aggregate output (i.e., $\alpha^* \tilde{R} + r$ in (2.28)) are affine linear increasing in the share α^* . Their relative effect $\frac{\alpha^* \tilde{c} + c^s}{\alpha^* \tilde{R} + r}$ is increasing in M^* if $\frac{c^r}{c^s} > \frac{\mathbb{E}(R)}{r}$. In addition to this, we have the effect of the fixed costs: The equilibrium costs in the banking sector rise with M^* through the two direct fixed costs of financial innovations d and f . Thus, the mix of the financial products matters for the size of the banking sector as follows: First, the size of the banking sector increases under the sufficient condition $\frac{c^r}{c^s} > \frac{\mathbb{E}(R)}{r}$ with the number N^* of financial products based on real underlyings because the fixed cost d -component is a convex increasing function of N^* . Second, the effect of a rise in the derivative multiplier ν^* on the size of the banking sector depends on its effect on the fixed costs relative to its effect on aggregate output (i.e., $\frac{f(1+\nu^*)N^* / (wI^*)}{\alpha^* \tilde{R} + r}$ in (2.28)). The sign of the relative effect depends on the level of the correlation neglect ϵ : If ϵ is not too large (i.e., if the bias due to ϵ in the portfolio choice α^* and thus on aggregate output is not too pronounced), the effect of an increased ν^* on the fixed costs outweighs the one on aggregate output. Then, the size of the banking sector increases with ν^* . The sufficient condition for this is $\epsilon = 0$ in addition to $\frac{c^r}{c^s} > \frac{\mathbb{E}(R)}{r}$.

λ^* is increasing in M^* if the extra costs implied by providing new financial products exceed relatively their gains on expected return. Under this, the model can explain a growing financial sector as a co-movement to the immense increase in the number of financial products. This is in line with the financialization over the last decades (see, e.g., Epstein (2005), Falkinger et al. (2015), Greenwood and Scharfstein (2013) or Philippon and Reshef (2012) for a discussion of financialization).

Proposition 2.4 summarizes the consequences of financial innovations on the aggregates:

Proposition 2.4. *Under the assumptions of Proposition 2.1 and Assumption 2.3, and irrespective of the presence of correlated derivatives and neglected correlation, aggregate output, its volatility and the size of the banking sector (the latter under the sufficient conditions $\frac{c^r}{c^s} > \frac{\mathbb{E}(R)}{r}$ and $\epsilon = 0$) depend, ceteris paribus, positively on the number of financial products.*

Proof. Main text. □

2.8 Conclusion

Since the recent crisis, the interconnections between the real economy and the financial sector attracted attention - within and outside academia. One aspect of the recent developments within the financial sector is the proliferation of the number of financial products. In a theoretical model in which banks compete for diversification-seeking households, I identify fundamentals and wrong beliefs as possible drivers of new financial products. Banks offer financial products to households for a fee. In the benchmark model the financial products are based on real investment projects. They generate stochastic returns and provide diversification possibilities. A financial innovation occurs whenever a bank offers a new financial product with an independent returns. The banking sector is in equilibrium if the number of financial products and the charged fees are consistent with optimal portfolio choice of the households and zero profit for banks (net of all cost, including management, issuing and advertising expenses). The model design allows for explicit solutions.

In the benchmark model, it is derived that in equilibrium more financial products are stimulated by changes in the fundamentals like increased volume of savings, more attractive financial assets and cost-reductions in the banking sector. Furthermore, by considering correlated financial products and wrong beliefs possible distortions are analyzed: Wrong beliefs about correlations bias households' portfolio choice. This leads to excessive risk taking and exaggerates financial innovations by correlated financial products. In particular, erroneous diversification assessment by households, reflected in a

neglect of correlation, induces banks to provide correlated derivatives. Thus, new financial products are not always the result of changes in fundamentals. They can also be induced by investors' neglected correlation and the subsequent cost-minimizing "cheating" behavior of banks. These results support policy recommendations that aim at enabling the households to clearly distinguish between financial products based on independent real projects and correlated derivatives so that the economy comes closer to an equilibrium in which the total number of financial products coincides with the number of financial products based on real underlyings.

The model allows the discussion of consequences of financial innovations: Expected aggregate output and the corresponding macroeconomic volatility depend positively on the number of financial products. Furthermore, the size of the banking sector, measured as the fraction of GDP spent for the provision of financial products and their management, increases with the number of financial products if the extra costs imposed by the provision of new financial products exceed relatively their gains on expected output. Such a rise of the banking sector size replicates qualitatively the co-movement of the financialization of the economy and the increase in the number of financial products.

3 Bank lending and firm dynamics in general equilibrium

Joint with Yingnan Zhao

3.1 Introduction

Access to financing is one of the main issues firms are dealing with. In general, financial constraints determine firms' development and their size distribution (Angelini and Generale, 2008). Especially for small and medium-sized firms with constrained access to bond or equity markets (The Economist, 2015), bank loans account for the primary part of external financing (Berger and Udell, 2002). This project analyzes how entrepreneurs and banks interact by modeling long-term credit relationships between them. Long-term credit relationships help to overcome information asymmetries through dynamic contracting. To the best of our knowledge we are the first who deal with such a long-term lending relationship in a general equilibrium framework which allows us to determine endogenously both the share of entrepreneurs as well as important aspects of firm dynamics such as size, growth and variance of growth of firms.

A key point of our model is the assumption of information asymmetry. This is, entrepreneurs have private knowledge about realized output levels of firms' production and banks cannot observe these. To deal with such repeated informational friction we take the paper of Smith and Wang (2006) on "dynamic credit relationships in general equilibrium" as a starting point. Like them we have banks and ex-ante identical households with finite life expectancy who either become entrepreneurs or workers. Workers supply labor, consume and save, whereas entrepreneurs run firms by hiring labor and capital. The realized output of firms is exposed to stochastic states of productivity. These are only observable to the entrepreneurs who report them to the banks. We extend Smith and Wang (2006) by adopting a production structure which allows for variable firm size like in Clementi and Hopenhayn (2006) (who work in a partial-equilibrium analysis). In particular, we use a technology with decreasing returns to scale. As in Clementi and Hopenhayn (2006) and Smith and Wang (2006) entrepreneurs finance production costs through loans from banks. Banks offer entrepreneurs long-term financial contracts, which determine

the optimal level of bank loans and state-contingent repayments. In recursive formulation these are determined together with future promised values as functions of today's promised values. A promised value is the continuation utility of an entrepreneur from consumption of future cash flows (net revenue generated from production by using bank loans minus repayments). The financial contracts are promise keeping and incentive compatible and fulfill the limited liabilities and the credibility constraints.

Our model structure allows us to determine the share of entrepreneurs endogenously and to see the effects of the dynamic lending-contracts on the size, growth, variance of growth of firms at different ages, and on the size distribution of firms in the economy in equilibrium. This extends Smith and Wang (2006) by the aspect of firm dynamics and it completes Clementi and Hopenhayn (2006) by the general equilibrium aspect. Further, it augments Dyrda (2014), Gross and Verani (2013) and Verani (2015) – who are also in a general equilibrium framework, but assume two types of households with different utility functions being either workers or entrepreneurs – with the endogenous determination of the share of entrepreneurs.

We calibrate our model; use it to get numerical results and to solve for the general equilibrium. Workers' saving and labor decision, entrepreneurs' choice of the optimal level of factor inputs and the optimal financial contract are derived numerically. We find that the optimal level of bank loans and state-contingent future promised values are increasing functions of today's promised values while the state-contingent repayments first increase and then decrease with the state variable. State-contingencies of future promised values and repayments are as follows: If entrepreneurs' report a high productivity state they are promised a higher future continuation utility, but they have to repay more today than if they report a low productivity state. This trade-off induces truth-telling about productivity realizations. By combining the three partial decision problems – of workers, entrepreneurs and banks, respectively – we close our model and determine the stationary general equilibrium. In general equilibrium our model predicts an interest rate of around 4%. This is a common number in literature. The share of entrepreneurs in our economy is found to be 8%, which corresponds approximately to the rate of self-employed in the U.S. (data from OECD). The firm dynamics resulting in general equilibrium from the optimal path of the promised values are as follows: There is a positive correlation between firm size and firm age. Furthermore, the growth of younger firms is on average larger and more volatile than that of older firms.

In addition to the numerical results and the economic explanation of them, we provide a discussion of technical issues which can cause problems in dynamic programming. These are, among others, starting value problems, extrapolation issues, sensitivity to parameter values and to functional forms as well as issues related to simulations.

The paper adds to the literature by modeling the long-term credit relationships between firms and banks in general equilibrium. Through dynamic contracting information asymmetries between banks and firms can be overcome. The analysis of repeated information asymmetries was initiated by Radner (1985) and Rogerson (1985). The dynamic programming approach to it with the recursive formulation of incentive compatible, optimal contracts was developed by Green (1987) and Spear and Srivastava (1987). Thomas and Worrall (1990), who extend the two-period, two-state problem of Townsend (1982) for any number of periods and finite state space, add to Green (1987) and Spear and Srivastava (1987) by focusing on the long-run asymptotic properties of the contracts. Such incentive compatible long-term contracts deliver on the one hand an insurance component if agents are exposed to idiosyncratic shocks which are unobservable (as in Green (1987), Thomas and Worrall (1990), Atkeson and Lucas (1992) or Atkeson and Lucas (1995)). On the other hand, they provide financing opportunities. In particular, contracts between risk-neutral banks and firms can support optimal lending- policies of banks which maximize the value of the firms (as in Quadrini (2004), Clementi and Hopenhayn (2006) or DeMarzo and Fishman (2007)).

The main contribution of this paper is the incorporation of dynamic financial contracts into a general equilibrium framework with an endogenous share of entrepreneurs and dynamic evolution of firms' size over age. Smith and Wang (2006), Dyrda (2014), Gross and Verani (2013) and Verani (2015) work in a similar model set-up. However, Smith and Wang (2006) do only consider projects with fixed units of capital and labor as input factors. They do not model an entrepreneur's optimal labor and capital decision under a more realistic production function. Hence, they cannot deal with firm dynamics. Dyrda (2014) does not consider a saving decision of workers by excluding them from the capital market. And in contrast to Dyrda (2014), Gross and Verani (2013) and Verani (2015), who do not incorporate firm entry, we determine the share of entrepreneurs in the economy endogenously. Technically, more complex equilibrium conditions are considered, which raises computational challenges.

Our model exhibits financial frictions which are the result of the information asymmetry. More precisely, we have borrowing constraints (i.e., firms do not get the efficient level of banks loans) as an endogenous result of the incentive-compatible long-term lending relationship between borrowers and the lender. This is like in the literature discussed above. Yet, on contrast to other contributions to the literature with long-term contracts between firms and banks, our model does not connect financial frictions to the issue of collateral as it is done in Clementi and Hopenhayn (2006) or Verani (2015). Nor do we allow for the possibility of auditing like in Verani (2015) or Albuquerque and Hopenhayn (2004) in an environment with limited enforcement.

Directly connected to the (endogenous) borrowing constraints are the dynamics of firm development in our model.¹ As a result of the long-term relation between the banks and firms, we predict that older firms are on average larger and that they grow less but more stable. These results are in line with the predictions from the dynamic contract models in Clementi and Hopenhayn (2006), Dyrda (2014), Gross and Verani (2013) and Verani (2015). Furthermore, they are consistent with the empirical regularities of firm dynamics.² In our model, the firm size distribution is more dispersed for older than for younger firms because their history of productivity realizations is more heterogeneous. That (endogenous) borrowing constraints have an impact on the size distribution of firms is consistent with the findings of Angelini and Generale (2008) and Cabral and Mata (2003) who find that younger firms, which are financially constrained, have in fact different (more skewed) firm size distributions.

The structure of the paper is as follow: Section 3.2 introduces the theoretical model. Therein, the workers' and entrepreneurs' problems and the role of financial intermediaries is described. Further, the recursive formulation of the dynamic lending contracts is presented and some theoretical properties are discussed. Section 3.3 provides the aggregation and equilibrium conditions and defines the stationary, general equilibrium. In Section 3.4 the calibration of the model and numerical results are presented. Section 3.5 discusses issues connected with dynamic programming. Section 3.6 concludes.

3.2 Model

3.2.1 Model set-up

Consider an infinite time horizon model with finite life expectancy. A continuum of ex-ante identical households are born at the beginning of each period. A household survives at the end of the period with an exogenous probability. Right after birth a household decides to become a worker or an entrepreneur. We assume that this choice of occupation is irreversible over lifetime. A worker supplies labor, consumes and saves part of its income. An entrepreneur runs a firm which uses labor and capital as inputs and consumes entrepreneurial income (net revenue from production). In addition to the households, there are banks which act as financial intermediaries between workers and entrepreneurs. Namely, they take annuity deposits from workers and offer contracts with financing in the form of bank loans to the entrepreneurs for their production. We assume

¹For an overview of the effects of financial frictions in a dynamic contract set-up on aggregate fluctuations / business cycle fluctuations see the literature discussed in Dyrda (2014) and Verani (2015). For the effects of access to credit on international trade see Gross and Verani (2013).

²See, for example, Evans (1987) or Hall (1987) for empirical literature on firm dynamics.

that banks are competitive so that they make zero profit in expectation from any financial contract they sign with entrepreneurs.

3.2.2 Households

Households are endowed with one unit of labor each period and no wealth at birth. The instantaneous utility function of the households (both workers and entrepreneurs) is $U(c, l)$, where c is the consumption level and l is the labor supply. $U(c, l)$ is decreasing in l and increasing, strictly concave and bounded in c . Households discount future with rate β .

The exogenous survival probability is Δ . We assume that the mass of newborns in each period is $1 - \Delta$, so that the mass of population is constant at 1. The share of the households in cohort τ who become entrepreneurs is λ_τ . λ_τ will be determined endogenously in equilibrium by the labor market clearing condition (see 3.27). Figure 3.1 summarizes the compositions of different cohorts' population size and their occupations at time t . It shows that at each point in time we have a distribution of workers and of entrepreneurs of different ages in the population.

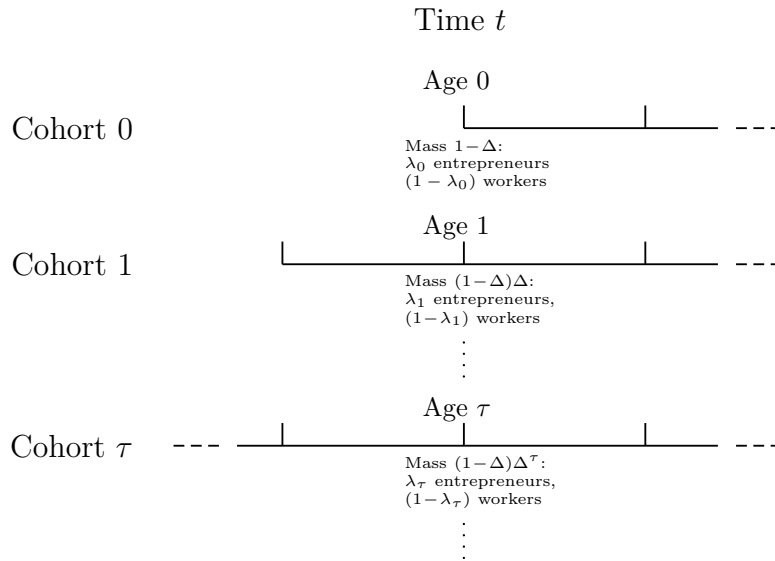


Figure 3.1: Different cohorts with infinite time horizon

3.2.2.1 Workers

In each period, workers supply labor for production and get wage income in return. Wage income as well as wealth can be used for consumption of final goods or as savings for wealth (and thus consumption) in future periods in the form of one-period annuity deposits in

the banks. In each period t , the workers of age τ buy at the end of the period $A_{\tau,t+1} \geq 0$ units of the annuity at price p_t^A . This entitles the worker to receive wealth level $A_{\tau,t+1}$ in period $t + 1$ conditional on survival. The annuity deposits are priced competitively (actuarially fair) such that banks make zero profit from offering them to the workers. This means, the aggregate amount of money received by the banks from workers plus the interest it generates within a period must be equal to what they give out in the next period. Formally, at time t ,

$$\sum_{\tau=0}^{\infty} (1 + r_{t+1})(1 - \Delta)\Delta^\tau p_t^A A_{\tau,t+1} = \sum_{\tau=0}^{\infty} (1 - \Delta)\Delta^{\tau+1} A_{\tau,t+1},$$

where $(1 - \Delta)\Delta^\tau p_t^A A_{\tau,t+1}$ is the aggregate payments of the workers of age τ at time t to buy the annuity. This generates an interest with rate $1 + r_{t+1}$ in the next period. The aggregate amount is redistributed to all workers from last period who are still alive this period, which is Δ -times the original size $(1 - \Delta)\Delta^\tau$ of each cohort. Therefore, zero-profit for banks implies that

$$p_t^A = \frac{\Delta}{1 + r_{t+1}}, \quad (3.1)$$

whereby the market-clearing interest rate r_{t+1} is endogenously determined in equilibrium.

The workers' problem of choosing labor supply l , consumption c and savings in annuities A' in an optimal way can be formulated in the following recursive way with today's wealth $A \geq 0$ as state variable:

$$V^W(A; r, w) = \max_{c, l, A'} \left\{ U(c, l) + \Delta\beta V^W(A'; r', w') \right\}, \quad (3.2)$$

subject to

$$\begin{aligned} c + p^A A' &= wl + A, \\ c &\geq 0, \quad l \in [0, 1], \quad A' \geq 0. \end{aligned} \quad (3.3)$$

$V^W(A; r, w)$ is the worker's value function (i.e., continuation utility) given today's wealth level A , interest rate r and wage rate w (determined endogenously in equilibrium). p^A is given by (3.1). A prime indicates variables of tomorrow. $\Delta\beta$ captures discounting and the fact that the worker survives with probability Δ . Denote the policy function of optimal saving A' and labor choice l , respectively, by

$$A_{\tau,t+1} = g(A_{\tau,t}; r_{t+1}, w_t), \quad l_{\tau,t} = h(A_{\tau,t}; r_{t+1}, w_t). \quad (3.4)$$

3.2.2.2 Entrepreneurs

Entrepreneurs run firms. They supply entrepreneurial labor and derive utility from consumption of net revenue from production. Entrepreneurs and firms are associated for the whole lifetime. Namely, a newborn household who becomes entrepreneur opens a firm and runs the firm for the entire lifetime until death; then the firm exits the market. Thus, the firm's exit rate is exogenously given by the household's death rate $1 - \Delta$.

Firms produce in each period under uncertainty a single output (numéraire), which can either be consumed or be used as capital. In each period a fixed amount of entrepreneurial labor L^E is needed for setting up the production. The production requires capital k and labor from workers l . The production function takes the form:

$$Y(k_t, l_t) = \theta_t F(k_t, l_t),$$

where $F(\cdot)$ reflects the production technology that transforms capital and labor inputs into the final product. It exhibits decreasing returns to scale. We assume the function to be continuous and strictly concave.

The level of θ_t represents the productivity at time t . In each period t the productivity is subject to an idiosyncratic shock with state space $\mathcal{S} = \{1, 2, \dots, S\}$ and the corresponding realization of states $\theta_t \in \Theta = \{\theta_1, \theta_2, \dots, \theta_S\}$. The shock is i.i.d. over entrepreneurs and time. The probability distribution of the states is $\{\pi_s\}_{s \in \mathcal{S}}$ with $\sum_{s \in \mathcal{S}} \pi_s = 1$. Without loss of generality, let $\theta_i < \theta_j$ if $i < j$. At any time t , each firm has an entire history of productivity realizations $\theta_\tau^t = (\theta_{t-\tau}, \theta_{t-\tau+1}, \dots, \theta_t)$, where τ is the age of the firm and $t - \tau + i$, $i \in \{0, 1, \dots, \tau\}$ is the calendar time when the firm was of age i . Note that the heterogeneity among firms is characterized by the different histories of productivity realizations.

We assume that the realization of productivity shock is private information to the entrepreneur. This reflects the information asymmetry between entrepreneurs and banks.

Prior to production (i.e., before the idiosyncratic shock is realized), the entrepreneurs need to purchase capital and pay the workers. By assumption, the entrepreneurs are neither endowed with wealth nor do they accumulate wealth from their production revenues over lifetime. This means, self-financing of production is excluded. Hence, they need external financing. We restrict the source of financing to bank funding. Bank loans and repayments arise from a lifetime financial contract between the bank and the entrepreneur. More specifically, the financial contract entitles the entrepreneurs each period to some amount of bank loans b , which is used to cover the production costs, and some repayments m after production.³ For a given level of loans and factor prices, the en-

³A detailed characterization of the financial contract, which includes bank loans b as well as repayments

trepreneurs determine the optimal capital and labor employment by maximizing expected output. The decision problem is

$$\max_{k_t, l_t} \mathbb{E}(\theta_t) F(k_t, l_t) \quad (3.5)$$

subject to

$$w_t l_t + (r_t + \delta) k_t \leq b_t,$$

where $(r_t + \delta)$ are the user cost of capital with δ being the depreciation rate of capital. We define

$$R(b_t; r_t, w_t) \equiv F(k_t^*, l_t^*) \quad (3.6)$$

with $k_t^* = k(b_t; r_t, w_t)$, $l_t^* = l(b_t; r_t, w_t)$ being the solution to (3.5) at which the marginal rate of transformation correspond to the relative factor price of capital and labor. Notice that firms' labor costs include only wage payments to workers. The implicit assumption is that the entrepreneurs do not supply the entrepreneurial labor L^E in the labor market of workers. In what follows we denote the labor supply from workers as labor.

The entrepreneur's consumption c_t in each period is given by net revenue from production, which is gross production $\theta_t R(b_t)$ minus repayments to banks m_t :

$$c_t^E = \theta_t R(b_t; r_t, w_t) - m_t. \quad (3.7)$$

Therefore, the expected lifetime utility of an entrepreneur is given by

$$V_0^E = \sum_{t=0}^{\infty} (\beta \Delta)^t \mathbb{E} U(c_t^E, L^E), \quad (3.8)$$

where expectation is with respect to current period realization of productivity, θ_t , as well as the history of realizations captured in b_t and m_t (as derived in Section 3.2.4). According to the properties of the utility function, natural limits of V_0^E are given by V_{min}^E and V_{max}^E , where

$$V_{min}^E \equiv \lim_{c \rightarrow 0} \frac{1}{1 - \beta \Delta} U(c, L^E) \text{ and } V_{max}^E \equiv \lim_{c \rightarrow \infty} \frac{1}{1 - \beta \Delta} U(c, L^E). \quad (3.9)$$

Remember that entrepreneurs do not make intertemporal savings decisions by assumption. Therefore, for given terms of the financial contract, maximization of expected lifetime utility in (3.8) is equivalent to maximizing the expected production output as given by (3.5).

The continuation utility of an entrepreneur at time t can in recursive formulation be

m , can be found in Section 3.2.4.

written as:

$$V_t^E = \mathbb{E} \left[U(c_t^E, L^E) + \beta \Delta V_{t+1}^E \right] \quad (3.10)$$

Notice that V_{min}^E and V_{max}^E are also the upper and the lower bound of continuation utilities of the entrepreneurs, respectively.

3.2.3 Financial intermediaries

The banks in the economy serve the role as financial intermediaries between saving households and producing firms. Namely, they take annuity deposits from workers and offer financial contracts to entrepreneurs. Banks act also as holder of capital by acquiring equity E . Banks' equity is the accumulated retained earnings from net cash flows of bank loans and repayments (see Section 3.3.1.3 for more details on equity). Overall, banks invest the annuity deposits from workers and own banks' equity as the capital supply of the economy into the production of entrepreneurs.

Banks are risk neutral profit maximizers and discount future at the current interest rate. There is free entry into the banking sector. This means in equilibrium banks expect zero profits from each single lending contract and thus size and ownership of the banks do not matter. Without loss of generality, we assume the existence of a representative bank holding a portfolio of all financial contracts with the entrepreneurs of all ages τ and with all heterogeneous histories of productivity realizations θ_τ^t .

3.2.4 Dynamic lending contract

The credit relation between banks and entrepreneurs is characterized by a lifetime binding financial contract. More specifically, following the standard dynamic contracting model (e.g., Thomas and Worrall (1990), Atkeson and Lucas (1992)), each firm signs a lifetime contract with a bank. Banks offer each newborn entrepreneur a take-it-or-leave-it lifetime binding financial contract.

We assume that both banks and entrepreneurs are fully committed to the contract in all possible future contingencies.

In the dynamic financial contract problem in recursive form, the continuation utility of an entrepreneur from future consumption, V_t^E as defined in (3.10), can be used as state variable (given interest and wage rate). Following the terminology of the literature, we call V_t^E the promised value. This means that the banks promise a continuation utility to the entrepreneurs by committing themselves to the terms of contract that imply a sequence of future consumption flows which generate the promised value. Therefore, given a promised value $V_t^E = V^E(\theta^{t-1}; r_t, w_t)$ as state variable – which includes the entire history of productivity realizations of an entrepreneur until time $t - 1$, θ^{t-1} – the con-

tract consists of $\{b(V_t^E; r_t, w_t), m(V_t^E, \theta_t; r_t, w_t), V^E(V_t^E, \theta_t; r_t, w_t)\}$.⁴ The first two terms are the bank loans to the entrepreneur, $b(V_t^E; r_t, w_t)$, and the repayments from the entrepreneur to the bank, $m(V_t^E, \theta_t; r_t, w_t)$. Loans are advanced before production, whereas repayments are made after the realization and after entrepreneurs' report of the current period productivity. Hence, loans are only contingent on today's promised value, whereas repayments are a function of today's promised value and the reported productivity level including time θ_t . (Note that according to the revelation principle, any equilibrium outcome can be achieved by a truth-telling mechanism. In particular, by imposing incentive constraints we can guarantee that entrepreneurs always report the actual realization of productivity θ_t . Therefore, we focus only on truth-telling contracts.) The third term, $V^E(V_t^E, \theta_t; r_t, w_t)$, is the next period's promised value given today's promised value and the productivity realization θ_t . In other words, $V_{t+1}^E = V^E(V_t^E, \theta_t; r_t, w_t)$ is the transition function of the state variable which incorporates the whole history of productivity realizations of an entrepreneur.

Since firms with the same promised value of today are assigned the same terms of contract (independent of time t or age τ), V^E is the state variable. V^E can be used as an indicator of firms in the equilibrium analysis (see Section 3.3.1.2 for aggregation of entrepreneurs).

3.2.4.1 Optimal financial contract

For given interest rate and wage rate, (r, w) , the optimal contract can be determined by the following program written in recursive form with the promised value, $V^E \in [V_{min}^E, V_{max}^E]$, as state variable:

$$P(V^E; r, w) = \max_{b, \{m_s, V_s^E\}_{s \in \mathcal{S}}} -b + \sum_{s \in \mathcal{S}} \pi_s \left[m_s + \frac{\Delta}{1+r} P(V_s^E; r', w') \right] \quad (3.11)$$

subject to

$$V^E = \sum_{s \in \mathcal{S}} \pi_s [U(\theta_s R(b; r, w) - m_s, L^E) + \beta \Delta V_s^E], \quad (\text{PK})$$

$$U(\theta_i R(b; r, w) - m_i, L^E) + \beta \Delta V_i^E \geq U(\theta_j R(b; r, w) - m_j, L^E) + \beta \Delta V_j^E, \quad \forall i, j \in \mathcal{S}, \quad (\text{IC})$$

$$m_s \leq \theta_s R(b; r, w), \quad \forall s \in \mathcal{S}, \quad (\text{LL})$$

$$V_s^E \in [V_{min}^E, V_{max}^E]. \quad (\text{CC})$$

$P(V^E; r, w)$ is the bank's expected profit (value function) from a financial contract with state variable V^E given r and w . b denotes the level of bank loans, $\{m_s, V_s^E\}_{s \in \mathcal{S}}$ are state-

⁴See Appendix C.1 for a detailed structure of the timing in the dynamic financial contract.

contingent repayments and future promised values, respectively. $\frac{\Delta}{1+r}$ captures discounting and the fact that the entrepreneur survives with probability Δ . V_{min}^E and V_{max}^E are given in (3.9).

(PK) is the promise keeping constraint. It indicates that the terms of the contract must be such that the expected utility from today's cash flows plus future promised values fulfill the promised value V^E .

(IC) captures the incentive constraints. In general, there is no guarantee that the reported productivity realization corresponds to the actual one. Yet, the incentive constraints induce that the truth-telling reporting strategy (weakly) dominates all other possible reporting strategies of entrepreneurs in terms of their expected utility, and thus eliminates incentives to misreport. Formally, we define:

Definition 3.1. *A contract is incentive compatible if the terms of the contract are such that (IC) is fulfilled $\forall i, j \in \mathcal{S}$.*

The constraints (LL) stand for limited liability. We impose that the entrepreneurs are liable for repayments to the bank at most to the extent of the production revenue (i.e., realized productivity shock times the production level corresponding to the bank loan level). Hence, a contract is feasible if the terms of the contract are such that the entrepreneurs consume a non-negative amount of the final products after any productivity realization. This leads to the following definition of a feasible contract:

Definition 3.2. *A contract is feasible if (LL) is fulfilled for $\forall s \in \mathcal{S}$.*

The credibility constraint (CC) imposes that banks can only promise utility values that are achievable with non-negative finite cash flows; otherwise, the promised value can only be granted by violating (LL) sometime in the future or is never satisfiable, respectively. More precisely, (CC) captures that banks can never promise (i) less utility than achievable by non-negative consumption for all future periods or (ii) more utility than by infinite consumption for all future periods.

Formally, we define an optimal financial contract as follows:

Definition 3.3. *For a given path $\{r_t, w_t\}_{t=0}^\infty$, the optimal dynamic contract is a sequence of functions $\{b(V_t^E; r_t, w_t), m(V_t^E, \theta_t; r_t, w_t), V^E(V_t^E, \theta_t; r_t, w_t)\}_{t=0}^\infty$ that solves program (3.11).⁵*

⁵For notational simplicity we suppress from now r_t and w_t in the sequence of functions of the contract $\{b(V_t^E), m(V_t^E, \theta_t), V^E(V_t^E, \theta_t)\}_{t=0}^\infty$ and in the value function $P(V^E)$ whenever it is not misleading.

3.2.4.2 Theoretical properties

In this part, we show theoretical properties of financial contracts under program (3.11).⁶ We first discuss general results about incentive compatible contracts and the simplification of incentive constraints. Then, we come to the properties of the optimal contract. Proposition 3.1 and 3.2 and Lemma 3.1 and 3.2 follow the properties of optimal social insurance in Ljungqvist and Sargent (2000) which are based on Thomas and Worrall (1990).

The following proposition defines the necessary condition of an incentive compatible contract:

Proposition 3.1. *Let $\theta_s > \theta_{s-1}, \forall s \in \mathcal{S}$. An incentive compatible contract satisfies $m_s \geq m_{s-1}$ and $V_s^E \geq V_{s-1}^E$.*

Proof. See Appendix C.2.1.1. □

This implies that banks induce truth-telling behavior of entrepreneurs by postponing rewards for reporting high productivity realization. If productivity is high repayments are high, but the future promised value is high, too.

Define the incentive constraints for all $i, j \in \mathcal{S}$ as:

$$C_{i,j} \equiv U(\theta_i R(b) - m_i, L^E) + \beta \Delta V_i^E - U(\theta_j R(b) - m_j, L^E) - \beta \Delta V_j^E, \quad (3.12)$$

where i is the actual state and j is the reported state. Then, the set of incentive constraints can then be simplified with the following lemma.

Lemma 3.1. *If the local downward constraints, $C_{s,s-1} \geq 0$, and the local upward constraints, $C_{s,s+1} \geq 0$, hold for each $s \in \mathcal{S}$, then the global constraints $C_{i,j} \geq 0$ hold $\forall i, j \in \mathcal{S}$.*

Proof. See Appendix C.2.1.2. □

Suppose for the following lemma and Proposition 3.2 and 3.3 that $P(V^E)$ is strictly concave – a fact which is observed in the numerics.

Using this and Lemma 3.1, we get the following property of the optimal contract.

Lemma 3.2. *For strictly concave $P(V^E)$, for all states $s \in \mathcal{S}$, the optimal contract implies that the local downward constraints $C_{s,s-1} \geq 0$ always bind, whereas the local upward constraints $C_{s-1,s} \geq 0$ never bind for $m_s > m_{s-1}$.*

Proof. See Appendix C.2.1.3. □

⁶In the optimal dynamic contract, the path of factor prices r_t, w_t is taken as given. For notational simplicity we suppress from now on $\{r_t, w_t\}$ in the $R(b)$ function whenever it is not misleading.

In addition, the optimal contract has the property of risk sharing:

Proposition 3.2. *For strictly concave $P(V^E)$, both the entrepreneurs' utility and the banks' profits are non-decreasing with a higher productivity realization, that is: Under an optimal contract, for $\theta_i > \theta_j$*

$$U(\theta_i R(b) - m_i, L^E) + \beta \Delta V_i^E \geq U(\theta_j R(b) - m_j, L^E) + \beta \Delta V_j^E, \quad (3.13)$$

$$-b + m_i + \frac{\Delta}{1+r} P(V_i^E) \geq -b + m_j + \frac{\Delta}{1+r} P(V_j^E). \quad (3.14)$$

Proof. See Appendix C.2.1.4. □

We introduce the efficient level of loan banks b^* , which is implicitly determined by

$$\mathbb{E}(\theta) R'(b^*; w, r) = 1. \quad (3.15)$$

Notice that the efficient level of loan banks corresponds to the optimal firm size if banks were the firm owners. It reflects that marginal productivity equals marginal costs of one more unit of bank loans.

Suppose for Proposition 3.3 that there are only two states in the state space, $\mathcal{S} = \{l, h\}$ with $\theta_h > \theta_l$.

Proposition 3.3. *For strictly concave $P(V^E)$ and for $m_s > m_{s-1}$, the optimal level of bank loans from the contract is not larger than the efficient level.*

Proof. See Appendix C.2.1.5. □

Note that this implies endogenous borrowing constraints, which are also existent in the models of Clementi and Hopenhayn (2006), Dyrda (2014), Gross and Verani (2013) and Verani (2015).

3.3 Aggregation and general equilibrium

So far we have characterized the optimization problems of the agents in the economy. More specifically, for a given sequence of factor prices $\{r_t, w_t\}_{t=0}^{\infty}$ and the share of entrepreneurs of each cohort $\{\lambda_\tau\}_{\tau=0}^{\infty}$, we get: (i) the workers' optimal path of consumption, wealth accumulation and labor supply from (3.2); (ii) the entrepreneurs' optimal path of capital and labor employment from (3.5); and (iii) the banks' optimal path of terms of contract with loans, repayments and future promised values from (3.11). Given the technical complexities, for combining the three partial parts to get the general equilibrium we focus on the stationary case with constant factor prices $\{r, w\}$ and a constant share of

entrepreneurs λ . Specifically, we consider the age-dependent, time-independent supplies and demands of labor and capital, consumption of workers and entrepreneurs, bank loans and repayments. We sum the individual decisions over the cohorts of all ages in the economy to get the aggregate demand and supply of labor, capital and goods. This allows us in the end to write down the equilibrium conditions and define the general equilibrium. More precisely, the equilibrium is then the prices $\{r, w\}$ and the share of entrepreneurs λ such that goods, labor and capital markets clear and banks make zero profit (see Sections 3.3.2 and 3.3.3).

3.3.1 Aggregation

3.3.1.1 Aggregation of workers

By aggregating the optimal consumption, saving and labor decision over individual workers of all ages τ , we get total consumption C^W , total deposits D and total labor supply L^S :⁷

$$C^W(r, w) = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^\tau c(A_\tau, r, w) \quad (3.16)$$

$$D(r, w) = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^\tau p^A A_{\tau+1} = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^\tau p^A g(A_\tau, r, w) \quad (3.17)$$

$$L^S(r, w) = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^\tau l_\tau = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^\tau h(A_\tau, r, w) \quad (3.18)$$

where $c(\cdot)$ is given by the workers' budget constraint (3.3) and $g(\cdot)$ and $h(\cdot)$ are the worker's policy functions defined in (3.4). $(1 - \Delta) \Delta^\tau$ is the mass of households of age τ . Note that heterogeneity among workers comes only from age differences; within a cohort all workers are identical in their lifetime decisions.

3.3.1.2 Aggregation of entrepreneurs

Aggregating over all entrepreneurs is more complicated because they are heterogeneous in two dimensions: Age and history of productivity realizations. In other words, there are firms of different ages τ and firms of the same age τ differ in productivity history θ^τ due to the idiosyncratic shocks.

History of productivity realizations of an entrepreneur aged τ , $\theta^\tau \in \Theta^\tau$ maps into a promised value V^E by applying the transition function $V_s^E = V^E(V^E, \theta_s; r, w)$ recursively

⁷For now the measure of workers is supposed to be 1. The equilibrium share of workers $(1 - \lambda)$ will be determined through the equilibrium conditions as given in Section 3.3.2.

with starting value V_0^E .⁸ The distribution of θ^τ among entrepreneurs of age τ corresponds to a stationary distribution of promised values, denoted by $\Psi_\tau(V^E)$.

Promised values V^E are translated by the optimal financial contract into bank loans and repayments, $\{b(V^E; r, w), m(V^E, \theta_s; r, w)\}$. For given bank loans, follow the optimal capital and labor employment, $\{k^*(V^E; r, w), l^*(V^E; r, w)\}$ defined by the solution to (3.5) where it is used that $b(V^E; \cdot)$ is a function of V^E .

We can aggregate the totals of the bank loans B , capital K^D and labor demand L^D over all cohorts as follows:⁹

$$B(r, w) = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^\tau \int b(V^E; r, w) d\Psi_\tau(V^E), \quad (3.19)$$

$$K^D(r, w) = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^\tau \int k^*(V^E; r, w) d\Psi_\tau(V^E) \quad (3.20)$$

$$L^D(r, w) = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^\tau \int l^*(V^E; r, w) d\Psi_\tau(V^E) \quad (3.21)$$

Furthermore, following the arguments and notations, the aggregate expected repayments from the entrepreneurs of all ages τ to banks are given by:

$$M(r, w) = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^\tau \sum_{s \in S} \pi_s \int m(V^E, \theta_s; r, w) d\Psi_\tau(V^E) \quad (3.22)$$

The expected aggregate output Y and the consumption of the entrepreneurs C^E are similarly given by:

$$Y(r, w) = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^\tau \sum_{s \in S} \pi_s \int \theta_s R(b(V^E; r, w); r, w) d\Psi_\tau(V^E), \quad (3.23)$$

$$C^E(r, w) = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^\tau \sum_{s \in S} \pi_s \int c(V^E, \theta_s; r, w) d\Psi_\tau(V^E), \quad (3.24)$$

where $R(\cdot)$ is defined in (3.6) and $c(\cdot)$ in (3.7).

3.3.1.3 Aggregation of banks' equity

Finally, banks' equity is the accumulated retained earnings from the flows of bank loans and repayments. In a stationary equilibrium, it is determined by

$$E(r, w) = (1 + r)E(r, w) + M(r, w) - B(r, w),$$

⁸The indifferent occupational choice condition, which must hold in equilibrium, requires that $V_0^E = V_0^W$ (see (3.26)).

⁹For now the measure of entrepreneurs is supposed to be 1. The equilibrium share of entrepreneurs λ will be determined through the equilibrium conditions as given in Section 3.3.2.

where $E(r, w)$ denotes the bank equity, $(1 + r)E(r, w)$ are the gross returns on previous equity and $M(r, w) - B(r, w)$ are the net aggregate payments from entrepreneurs.¹⁰ Rewriting the above equation, we have

$$E(r, w) = \frac{B(r, w) - M(r, w)}{r}. \quad (3.25)$$

3.3.2 Equilibrium conditions

In a stationary equilibrium (see Section 3.3.3 for the formal definition), there are simultaneously workers and entrepreneurs from all cohorts in the economy. Newborn households are indifferent with respect to their occupational choice. That is, the expected lifetime utility of becoming a worker is the same as that of becoming an entrepreneur. Formally, this means:

$$V_0^E = V^W(0; r, w), \quad (3.26)$$

where $V^W(0; r, w)$ is determined by program (3.2). Note that this equation defines the starting value of the state variable for the entrepreneurs, which is used in generating the life path of promised values in the numerical analysis by applying the optimal contracts (see Section 3.4.2).

In addition – as in standard general equilibrium theory – labor, capital and goods markets clear.

Labor market clearing requires that aggregate labor supply from workers equals aggregate demand for labor by the entrepreneurs. This is

$$\lambda L^D(r, w) = (1 - \lambda)L^S(r, w), \quad (3.27)$$

with $L^S(r, w)$ and $L^D(r, w)$ defined in (3.18) and (3.21), respectively, and λ being the endogenously determined share of the entrepreneurs in the economy.

Capital market clearing requires in equilibrium that capital supply in the economy, which consists of aggregate deposits from the workers plus banks' equity, is equal to capital demand:

$$K^S(r, w) \equiv (1 - \lambda)D(r, w) + \lambda E(r, w) = \lambda K^D(r, w), \quad (3.28)$$

where $D(r, w)$, $E(r, w)$ and $K^D(r, w)$ are given in (3.17), (3.25) and (3.20), respectively.

¹⁰Notice that this condition indicates that in the stationary equilibrium banks give on aggregate more loans than repayments they ask for; with the gap between B and M being exactly coverable by the interest from banks' equity. Thus, the level of equity is endogenously kept constant in the stationary case. Since we do not characterize the path of how the economy converges to the stationary equilibrium, we cannot show numerically how the accumulation of banks' equity converges to the stationary equilibrium level. However, we give in Appendix C.6 a non-rigorous intuition of how an economy may evolve from the very beginning of time to the stationary equilibrium.

The goods market is cleared if aggregate output equals the sum of households' consumption plus aggregate investments, where the latter is equal to depreciated capital in a stationary equilibrium. Formally, the condition is

$$\lambda Y(r, w) = (1 - \lambda)C^W(r, w) + \lambda C^E(r, w) + \delta \lambda K^D(r, w). \quad (3.29)$$

It is directly implied by the labor and the capital market clearing conditions, (3.27) and (3.28) as shown Appendix C.2.2.

Finally, banks' are assumed to make zero profit in expectation from each newly-signed contract in equilibrium. Under the indifferent occupational choice condition in (3.26), the zero-profit condition for banks is given by

$$P(V^W(0; r, w)) = 0. \quad (3.30)$$

3.3.3 Definition of general equilibrium

With the agents' optimal behavior derived from the respective optimization problems and the general equilibrium conditions, we can now define the stationary general equilibrium in the economy.

Definition 3.4. *A stationary general equilibrium is characterized by a stationary distribution of workers of different ages, and the corresponding capital and labor supply $\{A_\tau, l_\tau\}_{\tau=0}^\infty$, a stationary distribution of entrepreneurs of different ages, for each cohort a stationary distribution of promised values, $\{\Psi_\tau(V^E)\}_{\tau=0}^\infty$, and the corresponding capital and labor demand of the entrepreneurs, $\{k^*(V^E), l^*(V^E)\}$, bank loans and repayments of the banks, $\{b(V^E), m(V^E, \theta_s)\}_{s \in \mathcal{S}}$, and interest rate, wage rates and share of entrepreneurs, $\{r, w, \lambda\}$ such that for given (r, w) ,*

- (1) *workers maximize lifetime utility according to (3.2),*
- (2) *entrepreneurs maximize expected output according to (3.5),*
- (3) *banks offer profit-maximizing contracts subject to (PK), (IC), (LL), (CC) according to (3.11).*

The factor prices (r, w) and share of entrepreneurs λ are such that,

- (1) *labor, capital and goods market clear according to (3.27), (3.28) and (3.29).*
- (2) *banks make zero profit in expectation according to (3.30).*

We determine the stationary equilibrium numerically, but we do not deliver an analytical general proof for the existence of a stationary equilibrium.¹¹

3.4 Calibration and numerical results

Given the complexity of the problem, the stationary equilibrium is in the following determined numerically. To calibrate the model, we assume specific functional forms of the utility and the production function and give exogenous parameter values.

The households' utility function (workers and entrepreneurs) is given by

$$U(c, l) = -\exp(-\gamma c) - \eta l^2, \quad \gamma, \eta > 0. \quad (3.31)$$

It includes a CARA-part for consumption with γ being the absolute risk aversion and a parabola part for the disutility of labor supply. The form of the utility function gives us computational simplicity.

The production technology of the entrepreneurs exhibits decreasing return to scale:

$$Y(k, l) = \theta_s \bar{a} k^{\alpha_k} l^{\alpha_l}, \quad (3.32)$$

where θ_s denotes the state-dependent productivity realization, \bar{a} scales total factor productivity and α_k and α_l are the share of capital and labor, respectively. We simplify the state space \mathcal{S} to two states: “high” and “low” with productivity $\theta_h = \theta + \sigma$ and $\theta_l = \theta - \sigma$, $\sigma > 0$, and corresponding probability π_h and π_l , respectively.

For the exogenous parameters we take the values given in Table 3.1. The survival rate is chosen such that the death rate $1 - \Delta$ corresponds approximately to the empirical yearly exit rate of firms. The discount rate β is similar to standard values found in literature. Household preference parameter γ and η are internally calibrated such that workers' labor supply is about 30% of their labor endowment. Further, we assume both states are equally likely. Then, the values of θ_h and θ_l imply an expected productivity realization of $\theta = 1$, with standard deviation of 0.25. L^E corresponds to a third of an entrepreneur's labor endowment. α_k and α_l correspond to the usual capital and labor shares of output. The depreciation rate $\delta = 0.1$ corresponds to a common number in literature reflecting a quarterly depreciation rate of approximately 2.5%. The assumed utility function and the parameter values determine the boundaries of the promised value, $V_{min}^E = -9.26$ and $V_{max}^E = -0.49$ given by (3.9).

¹¹See Appendix C.4 for the numerical procedure and Appendix C.5 for a detailed description of the algorithm to find the stationary equilibrium numerically.

Table 3.1: Exogenous parameters

Parameters		Value
Survival rate	Δ	0.92
Discount rate	β	0.963
Household preferences	γ	2
	η	0.5
Probability of bad state	π_l	0.5
High productivity	θ_h	1.25
Low productivity	θ_l	0.75
Fixed entrepreneur labor	L^E	1/3
Share of capital	α_k	0.35
Share of labor	α_l	0.6
Productivity scale	\bar{a}	1/3
Depreciation rate	δ	0.1

3.4.1 Three optimization problems

We characterize first the numerical solutions to the three optimization problems. Specifically, for a given wage w and interest rate r , we solve for the workers' optimal consumption, saving and labor supply decision based on (3.2), the entrepreneurs' capital and labor demand as in (3.5) and especially the banks' optimal financial contract from (3.11).¹²

3.4.1.1 Workers' optimal decisions

Figure 3.2 depicts, as a function of the current period deposit wealth A , the workers' optimal consumption $c(A)$, the labor supply $l(A)$ and the saving decision $A'(A)$ corresponding to the policy functions given in (3.4) and the lifetime expected utility $V^W(A)$ for given w and r .¹³ They are in line with the results from standard lifetime utility maximization: Households consume more today and save more for tomorrow if their current wealth A is higher. One has $A'(A) > 0$ for all A , which means that households always decide to hold positive annuity deposits. Further, with more A they supply less labor because they are less dependent on labor income. Their lifetime expected utility, captured by the value function $V^W(A)$, is an increasing function in A , indicating that workers are better off if endowed with more wealth A .

¹²We use $w = 0.159965$ and $r = 0.04176$, which are the equilibrium values later determined numerically in the general equilibrium in Section 3.4.2 by using the search algorithm described in Appendix C.5. For simplicity we suppress w and r in the notation.

¹³See Appendix C.3.1 for the procedure to solve the recursive workers' problem numerically.

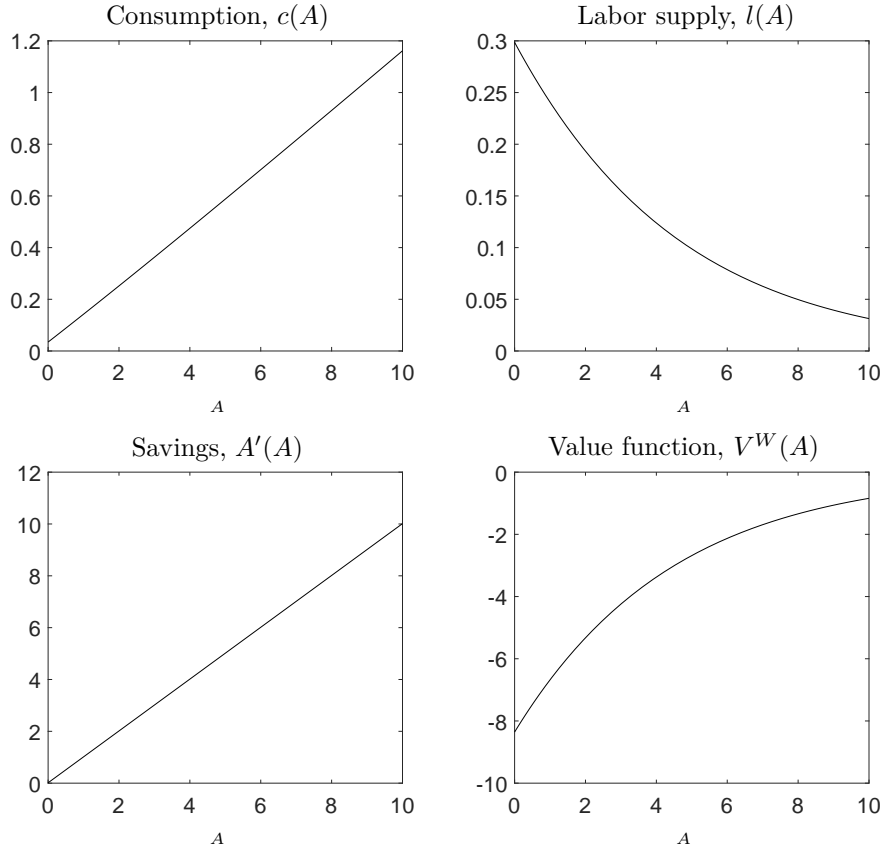


Figure 3.2: Solution to worker's problem

3.4.1.2 Entrepreneurs' optimal capital and labor employment

For a given level of bank loans b and factor prices w and r , the entrepreneur chooses optimally capital input and labor employment based on the decision problem in (3.5) as follows:

$$k^* = b \frac{1}{r + \delta} \frac{\alpha_k}{\alpha_k + \alpha_l} \quad \text{and} \quad l^* = b \frac{1}{w} \frac{\alpha_l}{\alpha_k + \alpha_l}. \quad (3.33)$$

Hence, $R(b) = \frac{1}{r+\delta} \frac{\alpha_k}{w} \frac{1}{\alpha_k + \alpha_l} \frac{\alpha_k^{\alpha_k} \alpha_l^{\alpha_l}}{(\alpha_k + \alpha_l)^{\alpha_k + \alpha_l}} b^{\alpha_k + \alpha_l} \bar{a}$.

Figure 3.3 shows this capital and labor demand of entrepreneurs as function of the bank loans b for given r and w . Capital and labor demand are linearly increasing functions in b . For the given form of the production function, the capital intensity is independent of the level of the bank loan b .

The outcomes indicate that the more bank loans firms get, the more input they are demanding. This implies that the size of production increases in the amount of available funds in the form of bank loans. Hence, bank loans determine the size of firms. Following this we will later use the amount of bank loans as the indicator of firm size and discuss

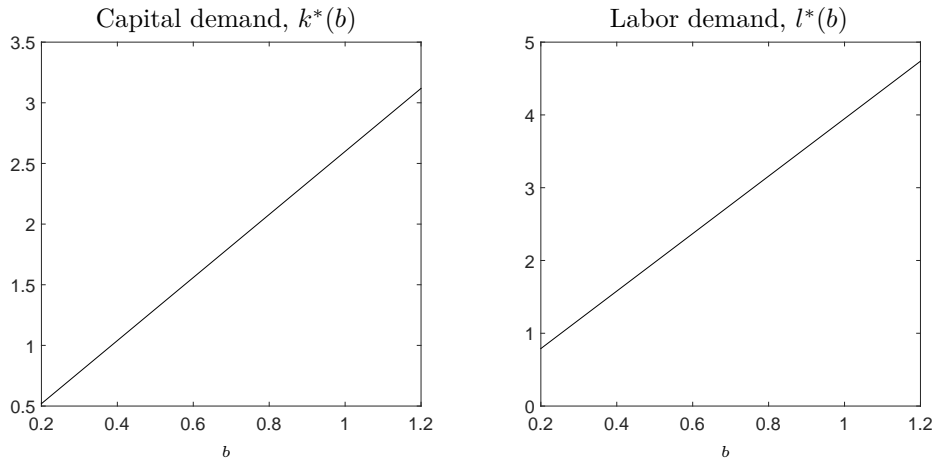


Figure 3.3: Solution to entrepreneur's problem

based on this the dynamics of average firm size, growth and variance of growth at different ages of the firms (see Section 3.4.4).

3.4.1.3 Banks' optimal financial contract

Figure 3.4 shows (for given r and w) as a function of today's promised value V^E (state variable), the banks' profit $P(V^E)$, state-contingent future promised value $V_s^E(V^E)$, state-contingent repayments $m_s(V^E)$ and the bank loans $b(V^E)$.¹⁴ Thereby, state-contingency is captured by the sub-indices with l standing for low and h for high productivity realizations.

The banks' profit is $P(V^E)$ is strictly concave. For V^E not close to V_{min}^E , $P(V^E)$ it is clearly decreasing in V^E .

The state-contingent future promised values, $V_l^E(V^E)$ and $V_h^E(V^E)$ are strictly increasing in V^E . Further, one can see from the subplot of $V_s^E(V^E)$ that $V_l^E < V_h^E$ and $V^E < V_h^E$. For values of V^E very close to V_{min}^E the lower credibility constraint (CC) is binding. In other words, without imposing the credibility constraint (CC), $V_l^E(V^E) < V_{min}^E$ would result for values of V^E very close to V_{min}^E , which contradicts $c \geq 0$ sometime in future.¹⁵

State-contingent repayments $m_s(V^E)$, $s \in \{h, l\}$ are non-monotonic; repayments $m_s(V^E)$ first increase in V^E and then decreases at higher promised values.¹⁶ The latter means, firms with a high promised value V^E have to repay less (even $m_s(V^E) < 0$) with the intuition that otherwise high V^E could not be realized without exploding V^E .

¹⁴See Appendix C.3.2 for the numerical procedure to solve the recursive formulated lending contract.

¹⁵This shows that accounting the credibility constraints is essential.

¹⁶Non-monotonicity can arise as a result of the functional forms of the utility, the production and the profit function, and their relative curvature compared to each other; the banks fulfill higher promised values V^E by both higher future promised utility and higher current consumption (through b cum m_s).

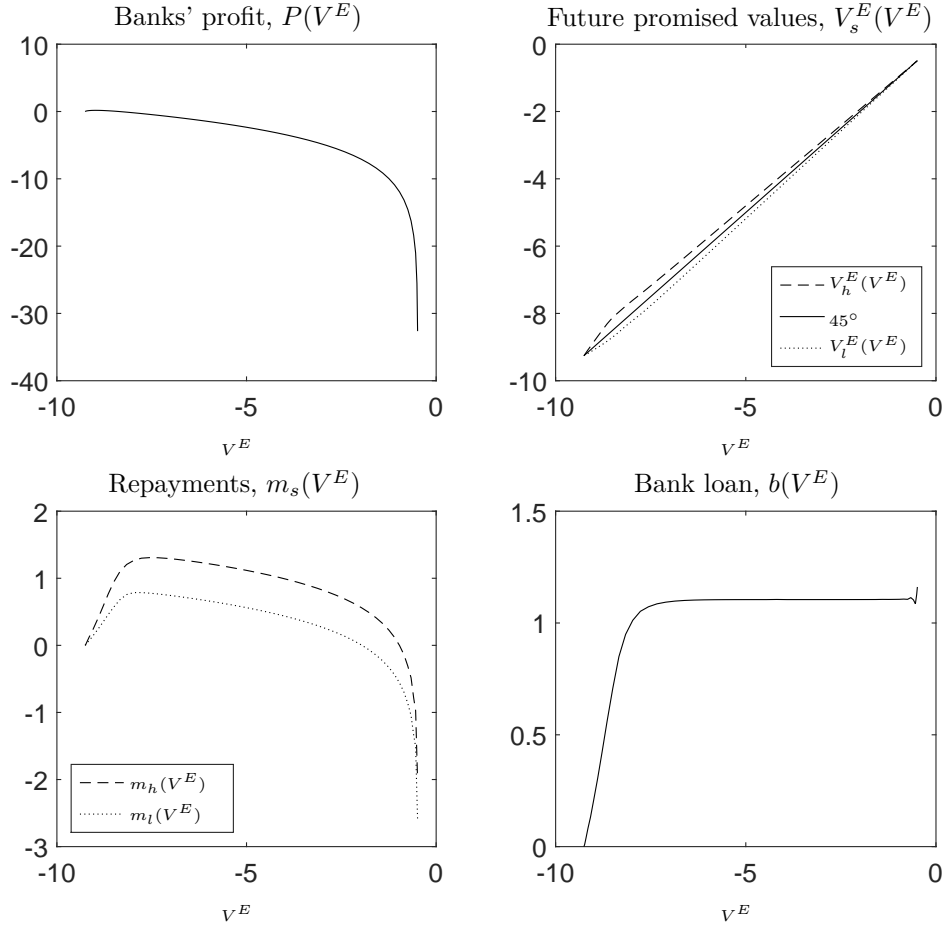


Figure 3.4: Optimal contract

path. Further, $m_l < m_h$ says that firms with a low productivity shock are spared from high repayments.

Note that the two subplots $V_s^E(V^E)$ and $m_s(V^E)$ for $s \in \{h, l\}$ reflect the theoretical results (see Proposition 3.1): Importantly, a postponed reward for reporting a high productivity state (high m_h , high V_h^E) and a postponed punishment for reporting a low productivity state (low m_l , low V_l^E) provide the entrepreneurs incentive to report the actual productivity realization.

Figure 3.4 shows further that the level of bank loan $b(V^E)$ is strictly increasing in V^E .¹⁷ By comparing the level of bank loans $b(V^E)$ with the expected repayment $\pi_l m_l(V^E) + (1 -$

¹⁷The specific shape of $b(V^E)$ is the result of the functional forms of the utility and the production function and their relative curvature compared to each other (see (C.9) in Appendix C.2.1). There are unstable $b(V^E)$ for V^E -values approaching V_{max}^E due to computational difficulties for values close to V_{max}^E . However, for determining the equilibrium this problem is negligible because firms hardly reach promised V^E -values in the region close to V_{max}^E when starting at $V_0^E = -8.36$ as derived in the general equilibrium (e.g., 65 years of always high productivity shock, which would leads to $V^E > -1$ has probability $(\Delta(1 - \pi_l))^{65} = 1.2 \cdot 10^{-22} \approx 0$). Further, the highest V^E reached by an entrepreneur in the simulation of our economy is only -1.74 .

$\pi_l)m_h(V^E)$ (see Appendix C.6), one see that for low V^E the expected repayments exceed the level of bank loans. Thus, banks retain earnings from the contracts at such state variable levels. For higher V^E the reverse holds which means that entrepreneurs retain deposited resources. For all V^E , $b(V^E)$ is smaller than the efficient level $b^* = 1.1814$ defined in (3.15); as stated in Proposition 3.3. $b(V^E)$ comes closer to the efficient level as V^E approaches V_{max}^E . The increasing function $b(V^E)$ means that firms with a higher promised value V^E get more bank loans and are thus larger. Hence, the transition function of the promised value, $V_s^E(V^E)$, is crucial in generating firm dynamics: For given current period productivity realization, the future promised value to entrepreneurs, $V_s^E(V^E)$, determines the level of tomorrow's bank loans and thus the evolution of the firm size. The relative level of bank loans available to a firm in two successive periods given by $\frac{b(V_s^E)}{b(V^E)}$, $s \in \{h, l\}$ depends on the productivity realization: A high productivity shock entitles the firm to more bank loans in the next period while a low productivity shock lowers b (see Figure 3.5, which gives a similar pattern as in Clementi and Hopenhayn (2006)).

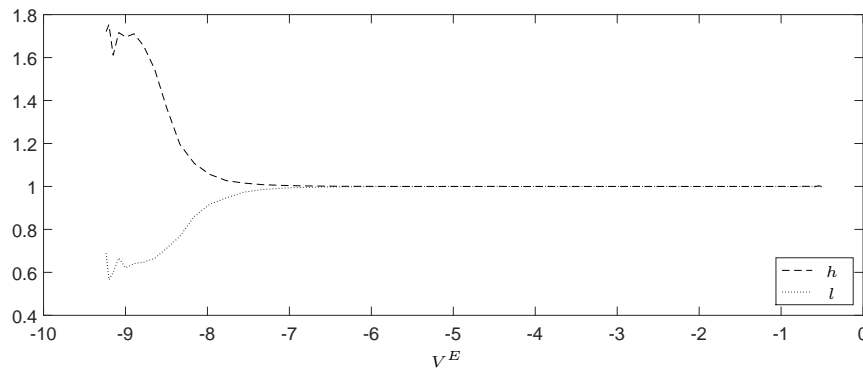


Figure 3.5: Relative change of bank loans, $\frac{b(V_s^E)}{b(V^E)}$, $s \in \{h, l\}$

3.4.2 General equilibrium

With the solutions of the three optimization problems, we can now determine the general equilibrium in our economy. The stationary equilibrium values of the endogenous factor prices (r, w) and the share of entrepreneurs λ are simultaneously found by labor and capital market clearing and banks' zero-profit condition. Thus, for determining the equilibrium, aggregate demands and aggregate supplies of labor and capital and the starting promised value $V^W(0; r, w)$ must be calculated.

The labor supply and part of the capital supply come from workers. In the stationary equilibrium, aggregating total deposits D and total labor supply L^S of all generations in the economy is computationally equivalent to the aggregation of deposits and labor

supply of one cohort over its lifetime as is implied by (3.17) and (3.18). The weights $(1 - \Delta)\Delta^\tau$ correspond then to the size of the cohort at the different ages τ . Since workers are homogeneous within cohorts with identical saving and working decision, it is numerically straightforward to compute the aggregate savings and labor supply using (3.17) and (3.18).¹⁸

To derive the demand for labor and capital we simulate life paths of entrepreneurs with stochastic shocks in their productivity and exogenous death.¹⁹ Thus, we have simulated the length of life for each entrepreneur and its history of productivity realizations, $\theta^t = \{\theta_1, \theta_2, \dots, \theta_t\}$. Starting at promised value $V_0^E = V^W(0, r, w)$, the simulated history of productivity realizations generates then for each entrepreneurs a lifetime sequence of promised values $\{V_1^E, V_2^E, \dots, V_t^E\}$ by applying the transition function $V_s^E(V^E)$ recursively.²⁰ To the sequence of $\{V_i^E\}_{i=1}^t$ correspond directly a sequence of repayments $\{m_i\}_{i=1}^t$ and a sequence of bank loans $\{b_i\}_{i=1}^t$. Aggregating these at t over all entrepreneurs we get in the end total repayments M and total bank loans B . The latter determines total labor and capital demands L^D and K^D according to (3.33): Because of linearity, (3.20) and (3.21), coincide with $K^D = B \frac{1}{r+\delta} \frac{\alpha_k}{\alpha_k + \alpha_l}$ and $L^D = B \frac{1}{w} \frac{\alpha_l}{\alpha_k + \alpha_l}$, respectively.

To determine the general equilibrium, we now use these aggregate demands and supplies. The share of entrepreneurs λ is determined by the labor market clearing condition (3.27). Namely, $\lambda = \frac{L^S}{L^D + L^S}$. The factor prices w and r are simultaneously determined by the capital market clearing condition (3.28) and the expected zero profit condition (3.30).²¹ The resulting equilibrium values of r, w and λ are shown in Table 3.2.

Table 3.2: Equilibrium parameter

Parameters		Value
Interest rate	r^*	0.04176
Wage	w^*	0.159965
Share of entrepreneurs	λ^*	0.07679

¹⁸See step d) in Appendix C.4 for the aggregation of the supply side.

¹⁹See steps a) and e) in Appendix C.4 for the description of the simulation procedure with $N^E = 10,000,000$ life paths.

²⁰See Figure C.4-C.6 in Appendix C.7.1 for three different examples of life paths of a 50 year old entrepreneur: The life path I illustrates a lucky life with many high productivity shocks. Life path II represents a life with a relatively balanced history of productivity realizations and life path III was driven by bad luck with many low productivity shocks. One can see that high productivity shocks tend to increase V^E overtime, while low productivity shocks lower it. The transition of V^E translates directly into the evolution of b and m .

²¹See step g) in Appendix C.4 for the procedure to determine the equilibrium in which the two conditions are jointly fulfilled and Appendix C.5 for the detailed description of the algorithm to find the stationary equilibrium numerically. We approximate the labor market clearing up to a residual of magnitude 0, the residual in the capital market is -0.0036 and the deviation from the zero-profit condition is -0.0030.

Our equilibrium interest rate is around 4%, which is a common number in literature. The equilibrium share of entrepreneurs λ is 7.7% and corresponds approximately to the rate of self-employed labor of around 7% in the U.S. over the last years (data from OECD).

The equilibrium lifetime expected utility and other equilibrium values are given in Table 3.3.

Table 3.3: Equilibrium values

Parameters		Value
Lifetime utility	$V_0^E = V^W(0, r^*, w^*)$	-8.3569
Total labor supply	L^S	0.2880
Total capital supply	D	0.1649
Total bank loans	B	0.8769
Total labor demand	L^D	3.4622
Total capital demand	K	2.2790
Total repayments	M	0.8626

The lifetime utility of entrepreneurs and workers is $V_0^E = V^W(0, r^*, w^*) = -8.36$. The total labor supply L^S corresponds to about a third of a worker's labor endowment, which is in line with standard values from the empirics. Further, from the amount of bank loans B and repayments M given in Table 3.3 we can calculate the amount of banks' equity E using (3.25). This indicates is an equity ratio $E/K = 15.1\%$. This number is above current levels of large international banks, but below the proposed level of 20% by Admati and Hellwig (2013).

3.4.3 Firm distributions

In this equilibrium, we can derive distributions for firm characteristics from the simulation of the paths of the entrepreneurs' lives. Figure 3.6 shows the distribution of entrepreneurs in the economy with respect to different characteristics: Age, promised values, repayments and bank loans.

Subplot (a) shows the distribution of entrepreneurs' ages. With a share of $1 - \Delta = 8\%$ most entrepreneurs are newborns. Then, one-year old represent a share of $(1 - \Delta)\Delta = 7.36\%$ and so on. Finally, the share of entrepreneurs older than 50 years account for only 0.12% in our economy.

Subplot (b) shows the distribution of promised values V^E . We get the histogram of the distribution of firm promised values $\Psi(V^E)$ as shown in Subplot (b) by counting the number of entrepreneurs in the economy in different bins of $V^E \in [V_{min}, V_{max}]$. The plot indicates clearly that the mass of the promised values lies around the starting value $V_0^E = -8.36$. Firm heterogeneity then arises from the different length and composition

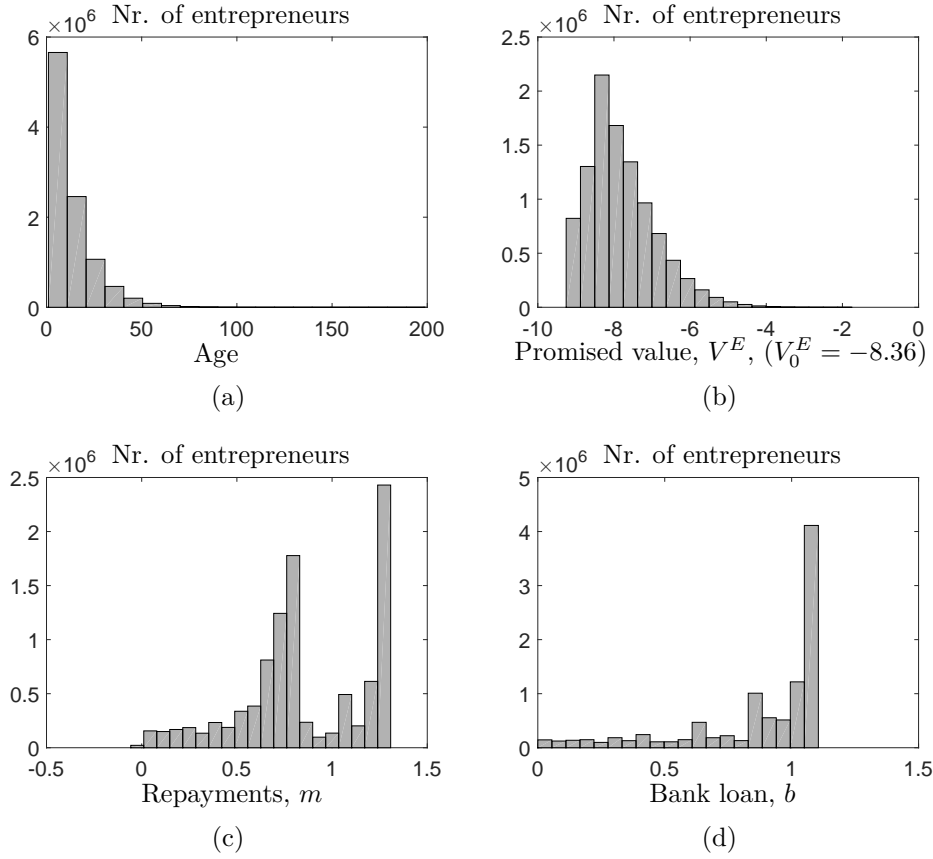


Figure 3.6: Distribution of age, promised values, repayments and bank loans

of productivity realizations over firms' lifetime. The further away from the starting value V_0^E , the lower is the density of V^E because longer and more heterogeneous life paths underlie such values.

Subplot (c) shows the distribution of repayments. It follows directly from the distribution $\Psi(V^E)$ (because V^E is the underlying state variable). Depending heavily on the current period productivity realization, the levels of repayments are separated into two groups. This means, the repayments exhibit two distinct sub-distributions because the difference in repayments of high and low state are relatively large (compare $m_h(V^E)$ and $m_l(V^E)$ in Figure 3.4).

Subplot (d) shows the distribution of bank loans. It also follows directly from the distribution $\Psi(V^E)$ (because V^E is the underlying state variable). It captures the firm size distribution measured by the levels of bank loans. From Figure 3.4 follow that for many V^E the optimal level of banks loans lies around the value $b(V^E) \approx 1.11$ (see relatively flat part in Figure 3.4). This means, many firms get such levels of banks loans so that the mode of the distribution of bank loans lies around this value. Thus, the negative skewness in the distribution of b is the result of the less strongly increasing part of $b(V^E)$ seen in Figure 3.4.

3.4.4 Firm dynamics

By considering now firm distribution of different cohorts separately (i.e., all entrepreneurs of the same age τ), the model allows us to get firm dynamics: Average firms' size, growth and variance of growth at different ages.

First, using the simulation of life path of entrepreneurs in Section 3.4.2 we generate the distribution of promised values $\Psi_\tau(V^E)$ of entrepreneurs at different ages.²² The development of $\Psi_\tau(V^E)$ for selected cohorts with age $\tau = \{0, 1, 2, 4, 10, 25, 120, 198\}$ is shown in Figure 3.7.

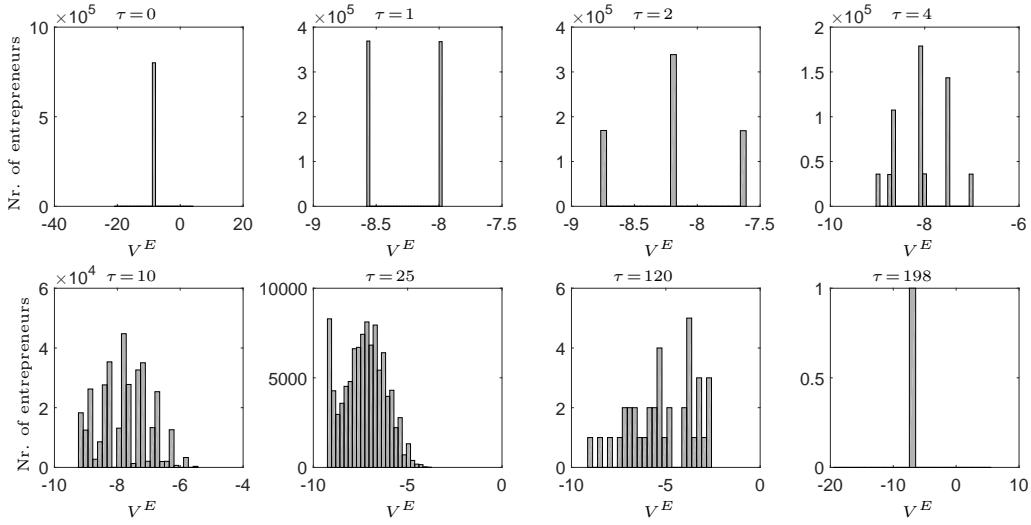


Figure 3.7: Development of entrepreneurs' promised value distributions

The newborns $\tau = 0$ are all identical with the same starting promised value $V_0^E = -8.36$. Surviving firms then experience either high or low productivity realizations and are updated with higher or lower future promised value levels, respectively. Over time, as histories of productivity realizations get more heterogeneous due to the i.i.d. shocks, the distribution $\Psi_\tau(V^E)$ for larger τ gets more dispersed. In addition, as age advances cohort size becomes smaller because firms have been exiting with the exogenous death rate $1 - \Delta$. Eventually, (almost) all firms of a given cohort exit the market so that the distribution $\Psi_\tau(V^E)$ of old cohorts consist of very few individual observations.

Following the cohort distribution, $\{\Psi_\tau(V^E)\}_{\tau=0}^\infty$, we can get firm dynamics such as average size, growth and variance of growth at different ages τ of entrepreneurs. Such firm dynamics are shown in Figure 3.8.²³

²²This maps directly into the distributions of bank loans and repayments. The corresponding distributions of b and m are shown in Figure C.7 and C.8 in Appendix C.7.2, respectively.

²³There is a decrease in firm size between the newborns and the one-year olds because the starting promised value, $V_0^E = -8.36$, is at the right end of the steep part of the $b(V^E)$ -function. Therefore, a

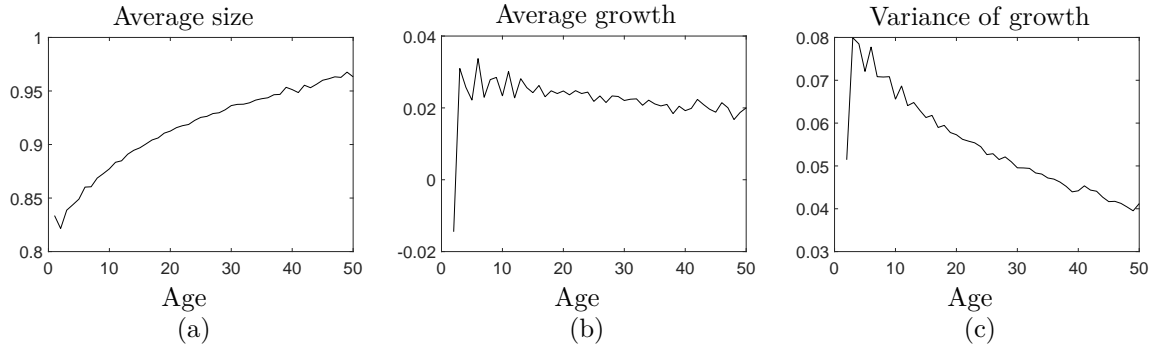


Figure 3.8: Firm dynamics

Figure 3.8 shows in Subplot (a) an increasing average size of firms at different ages. Firm size is measured in terms of the level of banks loans. Hence, our model predicts a positive relation between firm size and their age, which is in line with empirical observations.

In (b) we plot firms' average growth rates at different ages. We define the growth rate of a firm at age τ by the percentage change in bank loans relative to last period's loan, $g_\tau \equiv \frac{b_\tau - b_{\tau-1}}{b_{\tau-1}}$, where b_τ and $b_{\tau-1}$ are bank loans of today and of yesterday, respectively, of a firm with age τ . The average growth rate of all firms at age τ is measured by the mean of g_τ among all entrepreneurs in this cohort. The graph shows that firms' average growth is positive, but the rate decreases with firm age.²⁴ The same holds for the variance of the growth rate (i.e., the variance of g_τ) which is shown in Subplot (c). This means that on average older firms grow less, but in a more stable way.

These patterns are also found by Clementi and Hopenhayn (2006), Gross and Verani (2013) and Verani (2015), and are observed in industry data (e.g., Evans (1987)). This suggests that empirical firm dynamics can be explained by the design of the optimal financial contracts with endogenous borrowing constraints.

low productivity shock lowers b more than a high productivity shock increases b . In addition, since the history of productivity shocks is not very heterogeneous after one period (i.e., 50% are high and 50% are low) the decrease from the low productivity shock is directly reflected in the average size. For more periods the history of productivity shocks of entrepreneurs becomes more heterogeneous and the average is thus less dependent on single observations.

²⁴The observation discussed in the previous footnote is the reason for the outlier of the average growth (and also of the variance) in the first year. Note that the less smooth pattern for young firms comes from the fact that at the beginning firms have less different productivity paths, so that we have in this sense not enough cases of observations. The less smooth pattern for older firm arises since firms are dying and not many observations are left.

3.5 Discussion of dynamic programming

In this section, we discuss problems one could encounter in dynamic programming, mainly in solving the dynamic contract. They are, among others, starting value problems, extrapolation issues, sensitivities to functional forms and to parameter values, and issues related to the simulation.

3.5.1 Starting value problems

Dynamic programming problems are sensitive to initial guesses of value function and policy functions. In general, the convergence of dynamic programming algorithms is limited to a region close enough to the solution. This issue is especially obvious in the dynamic contract problem with relatively unconventional constraints. Therefore, our problem requires proper guesses of the starting values. We tried two ways: First, educated guess derived from the functional form of the utility function and, second, a “ground search”. For the educated guess, we consider a contract under perfect information with constant consumption for all periods and states. Then the value function, $P(V^E; r, w)$, which can serve as the starting value, is an affine linear transformation of the inverse of the utility function. In the ground search, we programmed a loop over a broad grid set of $\{b(V^E), m_h(V^E), m_l(V^E), V_h^E(V^E), V_l^E(V^E)\}$. We calculated for all combinations of the grid points the corresponding bank’s profits and then checked which of the combinations of the grid points maximize banks’ profits given that it fulfills all the constraints. These grid points are supposed to be somewhere in the region close to the solution of the optimal contract and can thus serve as the starting values. Overall, the initial guesses from the two ways are both good enough for solving the dynamic contract in our model. In the end, we used the first way to get the initial guesses as the second way requires relatively long computation time and a large amount of storing memory.²⁵

3.5.2 Extrapolation errors

In the numerical algorithm we generate a finite number of Chebychev grid points on the interval of the state variable.²⁶ Chebychev grid points have superior performances in function iterations in dynamic programming, yet, an extrapolation problem arises: The interval on which value function and policy functions are defined is larger than the range of

²⁵Note that even with only four grid points for each of the five choice variables there are already $4^5 = 1024$ combinations to be calculated and checked.

²⁶For Chebychev grid points we follow Judd (1998): The m grid points $\{x_k\}_{k=\{1, \dots, m\}}$ are set according to the coefficients of the Chebychev polynomial. We compute m Chebychev interpolation points $z_k = -\cos\left(\frac{2k-1}{2m}\pi\right)$ on $[-1, 1]$. Then we adjust it to our interval $[a, b]$, such that $x_k = (z_k + 1)\left(\frac{b-a}{2}\right) + a$.

the grid points, and thus extrapolation may be needed. With cubic splines interpolation, extrapolation close to the surroundings of the two grid boundaries is embedded in the code and thus performed automatically. However, the default extrapolation cannot guarantee that the image of the policy function remains in the domain of the state variable. To prevent this, we manually replaced the lowest Chebyshev point with the lower bound of the interval of the state variable before using it.

3.5.3 Sensitivity to parameter values and functional forms

Given the complexity of the dynamic contract problem, the numerical outcomes and the convergence of the iterations are sensitive to functional forms and parameter values. In our case, for example, with log-utility we would see a U -shaped $b(V^E)$, which may not necessarily lead to the same firm dynamics as we observe in Figure 3.8. Further, convergence of the problem is sensitive to the combinations of parameters. For example, not all combinations of (r, w) may guarantee convergence of the value function iteration when calculating the optimal dynamic contract. However, for combinations of (r, w) close to the equilibrium solution, the dynamic contract problem is in general stable; it always results in proper optimal contracts.

3.5.4 Simulation issues

To get the equilibrium, we grid-search (r, w) -combinations manually and check for the equilibrium conditions to hold. To have monotonicity in the aggregation variables of the simulation and comparability of different outcomes from the grid search, it is important that the simulation reflects a stationary distribution of life paths for all compared (r, w) -combinations. Otherwise, one cannot identify whether changes in the zero-profit condition and the capital market clearing condition come from the effect of updated (r', w') or from a changed combination of life paths. Furthermore, as is described in more detail in the algorithm in Appendix C.5, we use the observable fact that banks' profit and the excess demand in the capital market are both decreasing in r and w . In addition, the gap between the banks' profit and the excess demand in the capital market is decreasing in r and increasing in w . We use these signs and the (at least locally) observable monotonicity of the two conditions to restrict the region where the optimal equilibrium lies and get the direction for further searching.

3.6 Conclusion

This paper adds to the literature by modeling a dynamic credit relationship between banks and entrepreneurs in a general equilibrium model – which determines simultaneously the wage and interest rate and the share of entrepreneurs and which delivers firm dynamics. We have households, who decide, at the beginning of their life, to become either a worker or an entrepreneur. Workers supply labor and save in the form of annuity deposits. Entrepreneurs run firms and employ labor and capital for production. Productivity is stochastic and private knowledge to the entrepreneur. The production costs are financed with bank loans. To overcome the information asymmetry, loans and repayments are determined in long-term financial contracts with banks. More specifically, the financial contract between banks and entrepreneurs derived from a recursive formulation determines the optimal level of bank loan, state-contingent repayments and future promised values given today's promised values. The contracts are promise keeping and incentive compatible and fulfill limited liability and credibility constraints. In equilibrium banks make zero profit from the contract and the labor, the capital and the goods markets are cleared. The general equilibrium structure allows determining the wage, the interest rate and the share of entrepreneurs in equilibrium, as well as the size distribution of firms. Further, we get firm dynamics arising through the optimal financial contracts: The size of firms, measured by their level of bank loans, increases with the age of firms while their average growth and the variance of growth decreases with age.

Part III

Appendices

A Appendix: Chapter 1

A.1 Portfolio choice

Agent index l is skipped in the appendix. If financial intermediaries take ex-ante a fee in the form $T = p_{z_1}d + p_{z_2}(s - d)$, the expected utility maximization problem is given by:

$$\max_{s, \{f_\theta\}_{\theta \in \Theta}, d} \mathbb{E}U = \log(e_0 - \bar{e}_0) + \delta \left[\mu \sum_{\theta \in \Theta} \pi_\theta \log(e_\theta - \bar{e}_1) + (1 - \mu) \log(e_{\bar{\Theta}} - \bar{e}_1) \right]$$

subject to

$$e_0 + (1 + p_{z_2})s + (p_{z_1} - p_{z_2})d = y, \quad (\text{A.1})$$

$$e_\theta = \begin{cases} R_\theta f_\theta + rd, & \text{if } \theta \in \Theta \\ rd, & \text{otherwise} \end{cases} \quad (\text{A.2})$$

$$s = \sum_{\theta \in \Theta} f_\theta + d. \quad (\text{A.3})$$

Denoting by λ the Lagrange multiplier for constraint (A.3) the first-order conditions of the households' expected utility maximization problem give:

$$\frac{\partial \mathcal{L}}{\partial s} = -\frac{1 + p_{z_2}}{e_0 - \bar{e}_0} + \lambda = 0, \quad (\text{A.4})$$

$$\frac{\partial \mathcal{L}}{\partial f_\theta} = \delta \mu \pi_\theta \frac{R_\theta}{e_\theta - \bar{e}_1} - \lambda = 0, \quad (\text{A.5})$$

$$\frac{\partial \mathcal{L}}{\partial d} = -\frac{p_{z_1} - p_{z_2}}{e_0 - \bar{e}_0} + \delta \left[\mu \sum_{\theta \in \Theta} \pi_\theta \frac{r}{e_\theta - \bar{e}_1} + (1 - \mu) \frac{r}{rd - \bar{e}_1} \right] - \lambda = 0, \quad (\text{A.6})$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = s - \sum_{\theta \in \Theta} f_\theta - d = 0. \quad (\text{A.7})$$

Using (A.4), (A.5) and (A.6), we have

$$d = \frac{\delta(1 - \mu)}{\lambda \left(\frac{1 + p_{z_1}}{1 + p_{z_2}} - r/R \right)} + \frac{\bar{e}_1}{r}. \quad (\text{A.8})$$

where $R = \pi_\theta R_\theta$. From (A.2), (A.5) and (A.7), we have

$$s = \frac{\delta\mu}{\lambda} + (1 - r/R)d + \frac{1}{R}\bar{e}_1. \quad (\text{A.9})$$

In the end we have

$$\begin{aligned} d &= \frac{\delta(1-\mu)}{(1+\delta)P}(y - \bar{e}_0) + \frac{(1+\mu\delta)(1+p_{z_1}) - (1+\delta)(1+p_{z_2})r/R}{r(1+\delta)P}\bar{e}_1 \\ &= \frac{1-\mu}{1-p\rho} \frac{\delta}{1+\delta} \frac{y - \bar{y}}{1+p_{z_1}} + \frac{\bar{e}_1}{r}, \end{aligned} \quad (\text{A.10})$$

where $P \equiv (1+p_{z_1})(1-p\rho)$, $p \equiv \frac{1+p_{z_2}}{1+p_{z_1}}$, $\rho \equiv \frac{r}{R}$ and $\bar{y} \equiv \bar{e}_0 + \frac{\bar{e}_1(1+p_{z_1})}{r}$.

Combining (A.10) with (A.8) and solving for λ , we obtain

$$\frac{1}{\lambda} = \frac{y - \bar{y}}{(1+\delta)(1+p_{z_2})} \quad (\text{A.11})$$

Using this and (A.10) in (A.9), we have

$$\begin{aligned} s &= \frac{\delta}{(1+\delta)} \frac{y - \bar{y}}{1+p_{z_2}} \left[\mu + (1-\rho) \frac{p(1-\mu)}{1-p\rho} \right] + (1-\rho) \frac{\bar{e}_1}{r} + \frac{\bar{e}_1}{R} \\ &= \frac{\delta}{1+\delta} \frac{y - \bar{y}}{1+p_{z_2}} \frac{\mu - p\rho + p(1-\mu)}{1-p\rho} + \frac{\bar{e}_1}{r}, \end{aligned}$$

which can be rewritten in the form

$$s = \frac{\delta}{1+\delta} \frac{y - \bar{y}}{1+p_{z_2}} \left[1 + \frac{(p_{z_2} - p_{z_1})(1-\mu)}{(1+p_{z_1})(1-p\rho)} \right] + \frac{\bar{e}_1}{r}, \quad (\text{A.12})$$

where $p - 1 = \frac{p_{z_2} - p_{z_1}}{1+p_{z_1}}$ has been used.

Finally, (A.7), (A.10) and (A.12) give us

$$f \equiv \sum_{\theta \in \Theta} f_\theta = \frac{\mu - p\rho}{1-p\rho} \frac{\delta}{1+\delta} \frac{y - \bar{y}}{1+p_{z_2}} \quad (\text{A.13})$$

and from (A.1) we conclude

$$\begin{aligned} y - e_0 &= (1+p_{z_1})d + (1+p_{z_2})f \\ &= \frac{\delta}{1+\delta}(y - \bar{y}) + \frac{(1+p_{z_1})\bar{e}_1}{r}. \end{aligned} \quad (\text{A.14})$$

For the allocation of f on $f_\theta, \theta \in \Theta$, we combine (A.2) with (A.5) to get

$$\begin{aligned} f_\theta &= \pi_\theta \left[\frac{\delta\mu}{\lambda} + \frac{\bar{e}_1 - rd}{R} \right] \\ &= \pi_\theta \frac{\delta}{1 + \delta} \frac{y - \bar{y}}{1 + p_{z_2}} \left[\mu - \rho \frac{1 - \mu}{1 - p\rho} p \right] = \pi_\theta f, \end{aligned}$$

where (A.10) and (A.11) have been used for the second equation.

A.2 Corner solution for securities demand

To account for the non-negativity constraint $f_\theta \geq 0$ we have to add $\sum_{\theta \in \Theta} \psi_\theta f_\theta$ to the Lagrange function for max EU – with $\psi_\theta \geq 0$ denoting the Lagrange multiplier for $f_\theta \geq 0$. Then, the first order condition for f_θ changes to

$$\delta\mu\pi_\theta \frac{R_\theta}{e_\theta - \bar{e}_1} - \lambda + \psi_\theta = 0 \quad (\text{A.15})$$

with $\psi_\theta f_\theta \leq 0$.

Suppose that $f_\theta = 0$ for all θ . Then $s = d$ and

$$\begin{aligned} e_0 - \bar{e}_0 &= y - \bar{e}_0 - (1 + p_{z_1})d \\ e_\theta - \bar{e}_1 &= rd - \bar{e}_1 \end{aligned} \quad (\text{A.16})$$

and the first-order conditions

$$\begin{aligned} (s) \quad \lambda &= \frac{1 + p_{z_2}}{e_0 - \bar{e}_0} \\ (d) \quad \delta \left[\mu \sum_{\theta \in S} \pi_\theta \frac{r}{e_\theta - \bar{e}_1} + (1 - \mu) \frac{r}{rd - \bar{e}_1} \right] &= \lambda + \frac{p_{z_1} - p_{z_2}}{e_0 - \bar{e}_0} \end{aligned} \quad (\text{A.17})$$

reduce to

$$\delta \frac{r}{rd - \bar{e}_1} = \frac{1 + p_{z_1}}{e_0 - \bar{e}_0}.$$

With (A.16) this solves to

$$d = \frac{1}{1 + \delta} \left[\frac{\delta(y - \bar{e}_0)}{1 + p_{z_1}} + \frac{\bar{e}_1}{r} \right]. \quad (\text{A.18})$$

Substituting the solution into (A.16) gives us

$$\begin{aligned} e_0 - \bar{e}_0 &= \frac{1}{1 + \delta} \left[y - \bar{e}_0 - \frac{(1 + p_{z_1})\bar{e}_1}{r} \right] \\ e_\theta - \bar{e}_1 &= \frac{\delta r}{(1 + \delta)} \left[\frac{y - \bar{e}_0}{1 + p_{z_1}} - \frac{\bar{e}_1}{r} \right]. \end{aligned} \quad (\text{A.19})$$

Using this in (A.15) we obtain: $\psi_\theta \geq 0$ if and only if

$$\mu\pi_\theta R_\theta \leq \frac{1 + p_{z_2}}{1 + p_{z_1}} r \quad (\text{A.20})$$

where $\lambda = \frac{1+p_{z_2}}{e_0 - \bar{e}_0}$ has been used from (A.17).

Since $\pi_\theta R_\theta = R$, (A.20) reduces to

$$\frac{1 + p_{z_1}}{1 + p_{z_2}} \mu R \leq r,$$

which is equivalent to $R\mu(1 + p_{z_1}) \leq (1 + p_{z_2})r$.

Hence non-negativity $f_\theta > 0$, $\theta \in \Theta$, requires

$$R\mu(1 + p_{z_1}) > (1 + p_{z_2})r. \quad (\text{A.21})$$

A.3 Further proofs

Proof of Fact 1.3. With (1.11) and (1.12) the condition $y^L = b_L w_L > \bar{y} = \bar{e}_0 + \frac{(1+p_z)\bar{e}_1}{r}$ takes the form

$$A_x \Gamma_x \omega^{-\alpha_x} \left[b_L - \frac{\bar{e}_1}{r A_{z_1} \Gamma_{z_1}} \omega^{\alpha_{z_1}} \right] > \bar{e}_0 + \frac{\bar{e}_1}{r}.$$

The left side of the equation declines in ω . Thus $y^L > \bar{y}$ requires

$$\omega < \omega_L^+ \left(A_x, A_{z_1}, b_L, \bar{e}_0, \frac{\bar{e}_1}{r} \right),$$

where ω_L^+ is determined by the equation:

$$b_L = \left(\bar{e}_0 + \frac{\bar{e}_1}{r} \right) \frac{\omega^{\alpha_x}}{A_x \Gamma_x} + \frac{\bar{e}_1}{r} \frac{\omega^{\alpha_{z_1}}}{A_{z_1} \Gamma_{z_1}}.$$

□

Proof of Lemma 1.1. a) Let $B_1 \equiv A_x \Gamma_x \frac{b_L \bar{L}}{N}$ and $B_2 \equiv \frac{A_x \Gamma_x}{A_z \Gamma_z}$. Using (1.26) and (1.12), we have

$$\bar{w} = B_1 \omega^{-\alpha_x} (1 + \omega k), \quad p_z = B_2 \omega^{\alpha_z - \alpha_x}.$$

Then $\bar{\eta}$ can be reformulated as

$$\bar{\eta} = \frac{\bar{w} - \bar{y}}{1 + p_z} = \frac{B_1 \omega^{-\alpha_x} (1 + \omega k) - \bar{e}_0}{1 + B_2 \omega^{\alpha_z - \alpha_x}} - \frac{\bar{e}_1}{r},$$

where (1.18) is used to substitute \bar{y} .

To get the shape of $\bar{\eta}$, first notice that

$$\text{sign} \frac{\partial \bar{\eta}(\omega)}{\partial \omega} = \text{sign} \frac{\partial G(\omega)}{\partial \omega},$$

where $G(\omega) \equiv \frac{B_1(1+\omega k) - \bar{e}_0 \omega^{\alpha_x}}{\omega^{\alpha_x} + B_2 \omega^{\alpha_z}}$. Differentiating $G(\omega)$ we have

$$\frac{\partial G(\omega)}{\partial \omega} = \frac{\mathcal{L}(\omega)}{(\omega^{\alpha_x} + B_2 \omega^{\alpha_z})^2},$$

where

$$\begin{aligned} \mathcal{L}(\omega) = & B_1 \omega^{\alpha_x} \left[k(1 - \alpha_x) - \frac{\alpha_x}{\omega} \right] + B_1 B_2 \omega^{\alpha_z} \left[k(1 - \alpha_z) - \frac{\alpha_z}{\omega} \right] \\ & + \bar{e}_0 B_2 (\alpha_z - \alpha_x) \omega^{\alpha_x + \alpha_z - 1}. \end{aligned}$$

We have $\frac{\partial G(\omega)}{\partial \omega} > 0$ if and only if $\mathcal{L}(\omega) > 0$. For $\alpha_x + \alpha_z > 1$, $\mathcal{L}(\omega)$ is an increasing function in ω . Moreover,

$$\lim_{\omega \rightarrow 0^+} \mathcal{L} = -\infty, \quad \lim_{\omega \rightarrow +\infty} \mathcal{L} = +\infty.$$

Therefore, there exists a unique $\underline{\omega}$ with $\mathcal{L}(\underline{\omega}) = 0$ and: $\frac{\partial \bar{\eta}(\omega)}{\partial \omega} \gtrless 0$ if and only if $\omega \gtrless \underline{\omega}$. A rise in k or \bar{e}_0 shifts $\mathcal{L}(\omega)$ upward so that $\underline{\omega}$ declines. The impacts of B_1 , B_2 (and thus of A_x , A_z , $\frac{b_L \bar{L}}{N}$) on $\underline{\omega}$ are ambiguous because $\kappa_x < k < \kappa_z$ imply $k(1 - \alpha_x) - \frac{\alpha_x}{\omega} > 0$ and $k(1 - \alpha_z) - \frac{\alpha_z}{\omega} < 0$.

b) We have

$$\bar{\eta} = \frac{A_x \Gamma_x \frac{b_L \bar{L}}{N} \omega^{-\alpha_x} (1 + \omega k) - \bar{e}_0}{1 + \frac{A_x \Gamma_x}{A_z \Gamma_z} \omega^{\alpha_z - \alpha_x}} - \frac{\bar{e}_1}{r}.$$

By eye inspection we get:

$$\bar{\eta} \left(\omega \left| A_{x,+}, A_{z,+}, k_{+}, \frac{b_L \bar{L}}{N_{+}}, \bar{e}_0, \frac{\bar{e}_1}{r_{-}} \right. \right)$$

□

Proof of Fact 1.6. According to (1.38), $Z^S = A_z b_L \bar{L} \frac{\gamma_z^{\alpha_z}}{\gamma_z - \gamma_x} \omega^{-\alpha_z} (k\omega - \gamma_x)$, where $\kappa_j = \frac{\gamma_j}{\omega}$ has been used from (1.9).

We have $\frac{\partial \omega^{-\alpha_z} (k\omega - \gamma_x)}{\partial \omega} = \omega^{-\alpha_z} \left[(1 - \alpha_z)k + \frac{\alpha_z \gamma_x}{\omega} \right]$. This term is positive and decreasing in ω . \square

A.4 Data survey years 1995-2009

Table A.1: Parameters survey years 1995-2009

Parameter	Data	Source	Description
\bar{L}	109m	CPS	# Low-skilled employees
\bar{H}	41.1m	CPS	# High-skilled employees
h^L	1755.6	CPS	Yearly hours of low-skilled
h^H	2025.3	CPS	Yearly hours of high-skilled
α_x	0.44	CPS	Output ela. of high-skilled in X
α_{z_1}	0.54	CPS	Output ela. of high-skilled in Z_1
α_{z_2}	0.79	CPS	Output ela. of high-skilled in Z_2
A_x	32.46	CPS	Technology level in X
PT_{65}	\$ 11,213	U.S. Bureau of the Census	Real poverty threshold <65
PT^{65}	\$ 10,080	U.S. Bureau of the Census	Real poverty threshold >65
$LEratio$	3.83	LE from World Bank	Old-age ratio
r^f	0.0151	Federal Reserve Bank of St.Louis	Real federal funds rate
A_{z_1}	141.93	Model calibration· A_x -growth	Technology level in Z_1
A_{z_2}	201.88	Model calibration· A_x -growth	Technology level in Z_2
δ	0.385	Model calibration	Discount rate
μ	0.740	Model calibration	Certainty measure

Notes: The table shows the averaged values for the time range of survey years $t \in \{1995, \dots, 2009\}$.

Averages of $\alpha_{j,t} = \frac{\kappa_{j,t} \omega_{j,t}}{1 + \kappa_{j,t} \omega_{j,t}}$ with $\kappa_{j,t} = \frac{h_{j,t}^H \bar{H}_{j,t}}{h_{j,t}^L \bar{L}_{j,t}}$ and $\omega_{j,t} = \frac{w_{j,t}^H}{w_{j,t}^L}$, $j \in \{x, z_1, z_2\}$, $h_t^H = h_{x,t}^H$ and

$h_t^L = h_{x,t}^L$. $A_{x,t} = \frac{w_{x,t}^L}{\Gamma_{x,t} \omega_{x,t}^{-\alpha_{x,t}}}$ with $\Gamma_{x,t} = \alpha_{x,t}^{\alpha_{x,t}} (1 - \alpha_{x,t})^{1 - \alpha_{x,t}}$. PT is the average, real poverty

threshold of a two-people household (nominal values are adjusted by using the CPI-U adjustment factor to 1999 dollars (i.e., for the base survey year 2000) from CPS with PT_{65} denoting the relevant value for households younger than 65 and PT^{65} denoting the value relevant for older ones. $LEratio$ is the average ratio of working-time to retirement: $(65 - 20)/(LE_t - 65)$, where LE_t denotes life expectancy in year t ; 65 is the retirement age and 20 is the assumed start of the working-life. r^f is the average, real effective federal funds rate (effective federal funds rate adjusted with the CPI-U adjustment factor from CPS). See bibliography for details on data sources.

B Appendix: Chapter 2

B.1 Derivations

B.1.1 Derivation of the equilibrium

Using (2.7) in (2.10), we have

$$\hat{\tau}(N) = \tilde{c}\alpha^* + c^s + \frac{dN^2 + fN}{I^*w}$$

From (2.6) we know

$$v^*(N, \hat{\tau}(N)) = \alpha^* \tilde{R} - \frac{\gamma \sigma_{PF}^2}{2} + r - \hat{\tau}(N),$$

where $\sigma_{PF}^2 = \alpha^{*2} \sigma_B^2$. Hence, the equilibrium condition $\frac{dv^*(N, \hat{\tau}(N))}{dN} = 0$ reads

$$\frac{\partial \alpha^*}{\partial N}(\tilde{R} - \tilde{c}) - \frac{\gamma}{2} \frac{\partial \sigma_{PF}^2}{\partial N} = \frac{2dN + f}{I^*w}. \quad (\text{B.1})$$

B.1.1.1 Derivation of N^*

For $\alpha^*(N) = N \frac{\tilde{R}}{\gamma \sigma^2}$ and $\sigma_B^2 = \frac{\sigma^2}{N}$, we have $\sigma_{PF}^2 = N \frac{\tilde{R}^2}{\gamma^2 \sigma^2}$ so that the respective derivatives are: $\frac{\partial \alpha^*}{\partial N} = \frac{\tilde{R}}{\gamma \sigma^2}$ and $\frac{\partial \sigma_{PF}^2}{\partial N} = \frac{\tilde{R}^2}{\gamma^2 \sigma^2}$. Thus, the equilibrium condition (B.1) reduces to $\frac{\tilde{R}}{\gamma \sigma^2}(\tilde{R} - \tilde{c}) - \frac{\tilde{R}^2}{2\gamma \sigma^2} = \frac{2dN + f}{I^*w}$. It solves to

$$N^* = \frac{I^*w \tilde{R}(\tilde{R} - \tilde{c}) - \frac{\tilde{R}^2 I^*w}{2} - f\gamma \sigma^2}{2d\gamma \sigma^2} \quad (\text{B.2})$$

and can be rewritten in the form of (2.12). The second order condition $\frac{d^2 v^*(N, \hat{\tau}(N))}{dN^2} = -\frac{2d}{I^*w}$ is negative. N^* gives the maximum and the unique equilibrium. N^* plugged into the zero-profit condition (i.e., $\hat{\tau}(N^*)$) gives τ^* as presented in equation (2.13).

B.1.1.2 Derivation of N_D^*

Use $\tilde{f} = (1 + \nu)f$ in (B.2) and rewrite it to get the form in (2.17).

B.1.1.3 Derivation of $N_{D,NC}^*$

For $\alpha^*(N, \nu, \epsilon) = \frac{(1+\nu)N}{1+\nu(1-\epsilon)} \frac{\tilde{R}}{\gamma\sigma^2}$ and $\sigma_B^2(N, \nu, \epsilon) = \frac{1+\nu(1-\epsilon)}{(1+\nu)N} \sigma^2$, we have $\sigma_{PF}^2 = \frac{(1+\nu)N}{1+\nu(1-\epsilon)} \frac{\tilde{R}^2}{\gamma^2\sigma^2}$ so that $\frac{\partial \alpha^*}{\partial N} = \frac{1+\nu}{1+\nu(1-\epsilon)} \frac{\tilde{R}}{\gamma\sigma^2}$ and $\frac{\partial \sigma_{PF}^2}{\partial N} = \frac{1+\nu}{1+\nu(1-\epsilon)} \frac{\tilde{R}^2}{\gamma^2\sigma^2}$. Thus, the equilibrium condition (B.1) reduces to $\frac{1+\nu}{1+\nu(1-\epsilon)} \frac{\tilde{R}}{\gamma\sigma^2} (\tilde{R} - \tilde{c}) - \frac{1+\nu}{1+\nu(1-\epsilon)} \frac{\tilde{R}^2}{2\gamma\sigma^2} = \frac{2dN+\tilde{f}}{I^*w}$, where $\tilde{f} = (1+\nu)f$. The condition solves to

$$N_{D,NC}^* = \frac{I^*w\tilde{R}(\tilde{R} - \tilde{c}) - \frac{\tilde{R}^2 I^*w}{2} - \tilde{f} \frac{1+\nu(1-\epsilon)}{1+\nu} \gamma\sigma^2}{2d \frac{1+\nu(1-\epsilon)}{1+\nu} \gamma\sigma^2}, \quad (\text{B.3})$$

which, by using $\tilde{f} = (1+\nu)f$, can be rewritten in the form given in (2.19).

B.1.2 Derivation of $\sigma_B^2(N, \nu, \epsilon)$

A package of N independent financial products has variance $\sigma_N^2 = \sigma^2/N$. There are $(1+\nu)$ of such packages with correlation $\rho = 1$. Households are diversification-seeking and thus put equal weight $1/(1+\nu)$ on each of the package (argumentation as in Lemma 2.1). The variance of the risky bundle B consisting of N and $D = \nu N$ assets is therefore (equally weighted sum of covariances):

$$\sigma_B^2(N, \nu) = \frac{1}{1+\nu} \sigma_N^2 + \frac{(1+\nu) - 1}{1+\nu} \rho \sigma_N^2$$

However, a share ϵ of ρ is neglected:

$$\begin{aligned} \sigma_B^2(N, \nu, \epsilon) &= \frac{1}{1+\nu} \sigma_N^2 + \frac{(1+\nu) - 1}{1+\nu} (\rho - \epsilon) \sigma_N^2 \\ &= \frac{1 + \nu(\rho - \epsilon)}{(1+\nu)N} \sigma^2 \end{aligned}$$

with $\rho = 1$

$$\sigma_B^2(N, \nu, \epsilon) = \frac{1 + \nu(1 - \epsilon)}{(1+\nu)N} \sigma^2.$$

Clearly, $\sigma^2(N, \nu, \epsilon) \rightarrow \frac{\sigma^2}{N}$ as $\epsilon \rightarrow 0$ and $\sigma^2(N, \nu, \epsilon) \rightarrow \frac{\sigma^2}{(1+\nu)N}$ as $\epsilon \rightarrow 1$.

B.1.3 Derivations of analytical results for endogenizing ν

B.1.3.1 Derivation of cost-minimizing ν^*

Minimizing and setting the first order condition (FOC) equal to zero: $\frac{\partial C_{tot}(\nu, \epsilon)}{\partial \nu} = 2da^2 \frac{(1-\epsilon)(1+\nu) - (1+\nu(1-\epsilon))}{(1+\nu)^2} \frac{1+\nu(1-\epsilon)}{1+\nu} + af(1-\epsilon) \stackrel{!}{=} 0 \Leftrightarrow \frac{f}{2da} = \epsilon \frac{1-\epsilon+\nu}{(1+\nu)^3}$ given in (2.22).

For the SOC follows: $\frac{\partial^2 C_{tot}(\nu, \epsilon)}{\partial \nu^2} = -2da^2 \epsilon \frac{(1-\epsilon)(1+\nu) - 3(1+\nu(1-\epsilon))}{(1+\nu)^4}$, which is positive for $\nu \geq 0$

and $0 < \epsilon < 1$. Thus, ν^* , which is implicitly determined in (2.22), is a minimum. It gives a unique, positive ν^* for $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$: Define the implicit function $F \equiv \epsilon \frac{1-\epsilon+\nu^*}{(1+\nu^*)^3} - \frac{f}{2da}$ from (2.22). There is a unique, positive ν^* for $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$ because $\frac{\partial F}{\partial \nu^*} < 0 \Leftrightarrow 3(\frac{1}{1-\epsilon} + \nu) > (1+\nu)$ for all $\epsilon \in [0, 1]$, $\lim_{\nu^* \rightarrow 0} F > 0$ for $\epsilon > \underline{\epsilon} = \frac{f}{2da+f}$ and $\lim_{\nu^* \rightarrow \infty} F < 0$ due to $\lim_{\nu^* \rightarrow \infty} \epsilon \frac{1-\epsilon+\nu^*}{(1+\nu^*)^3} = 0$ (i.e., F crosses the zero-axis for positive ν^* only once).

Graphically the cost-minimization means that the locus, which describes all (N, ν) -combination providing the same perceived level of diversification a (iso-diversification curve), and the iso-cost curve are tangent:

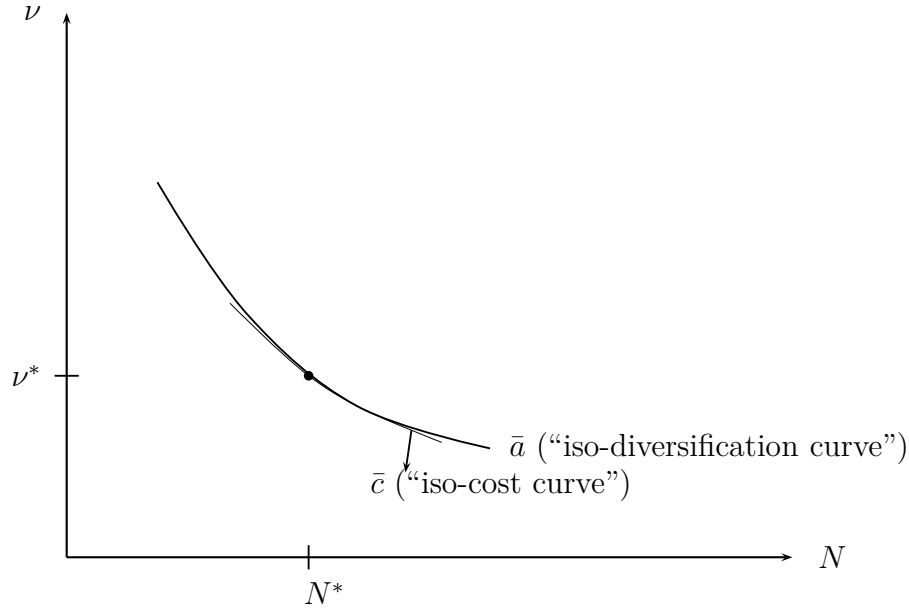


Figure B.1: Iso-diversification and iso-cost curves

Notes: The figure plots an iso-diversification curve for a perceived level of diversification \bar{a} : $\nu(\bar{a}, N) = \frac{\bar{a}-N}{N-\bar{a}(1-\epsilon)}$ with $\partial \nu(\bar{a}, N)/\partial N \leq 0$ and $\partial^2 \nu(\bar{a}, N)/\partial N^2 \geq 0$ because $N > \bar{a}(1-\epsilon)$. The iso-cost curve \bar{c} captures total costs $\bar{C}_{tot}(N, \nu, \epsilon)$ with $\partial \nu(\bar{c}, N)/\partial N \leq 0$ and $\partial^2 \nu(\bar{c}, N)/\partial N^2 \geq 0$ by implicit differentiation of $\bar{C}_{tot}(N, \nu, \epsilon)$.

B.1.3.2 Derivation of $\underline{\epsilon}$ and $\tilde{\epsilon}$

(2.22) holds only for $0 < \nu^* < \frac{a-1}{1-a(1-\epsilon)}$ (i.e., only if the constraints $\nu^* \geq 0$ and $\nu^* \leq \frac{a-1}{1-a(1-\epsilon)}$ (because of $N^* \geq 1$) are non-binding): Using $\nu_{\min}^* = 0$ in (2.22), we have $\frac{f}{2da} = \epsilon \frac{1-\epsilon+0}{(1+0)^3}$ such that $\underline{\epsilon} = \frac{f}{2da+f}$ with $\underline{\epsilon} \in (0, 1)$. Since ν^* in (2.22) rises in ϵ if $0 < \epsilon < 1$ (see below) this means that for all $\epsilon < \underline{\epsilon}$ we set $\nu^* = \nu_{\min}^* = 0$. Further, using $\nu^* = \frac{a-1}{1-a(1-\epsilon)}$ in (2.22), we have $\frac{f}{2da} = \tilde{\epsilon} \frac{1-\tilde{\epsilon}+\frac{a-1}{1-a(1-\tilde{\epsilon})}}{(1+\frac{a-1}{1-a(1-\tilde{\epsilon})})^3}$. This gives $\tilde{\epsilon}_{1,2} = \frac{4(a-1)d+af \pm \sqrt{f}\sqrt{8(a-1)d+a^2f}}{2a(2d+f)}$. Since there is only one positive solution ν^* determined by (2.22) and $\tilde{\epsilon}_1 = \frac{4(a-1)d+af - \sqrt{f}\sqrt{8(a-1)d+a^2f}}{2a(2d+f)} < \frac{4(a-1)d+af + \sqrt{f}\sqrt{8(a-1)d+a^2f}}{2a(2d+f)} = \tilde{\epsilon}_2$ with $0 < \tilde{\epsilon}_2 \leq 1$ for $a \geq 1$, $d \geq 0$ and $f \geq 0$ we have $\tilde{\epsilon} \equiv \tilde{\epsilon}_2$

(i.e., I can exclude $\tilde{\epsilon}_1$ because $\nu^* = \frac{a-1}{1-a(1-\tilde{\epsilon}_1)} < 0$). This implies that for all $\epsilon > \tilde{\epsilon}$ we set $\nu^* = \frac{a-1}{1-a(1-\epsilon)}$. Thus, $\nu^*_{max} = \frac{a-1}{1-a(1-\tilde{\epsilon})}$ and for $\epsilon = 1$ we have $\nu^* = a - 1$.

B.1.3.3 Derivation of comparative statics on ν^* and N^*

Use the differentiable implicit function F defined in Appendix B.1.3.1. The implicit function theorem gives $\nu^{*'}(\epsilon) = -\frac{\frac{\partial F}{\partial \epsilon}}{\frac{\partial F}{\partial \nu^*}} = -\frac{\frac{1+\nu^*-2\epsilon\nu^*+\nu^{*2}\epsilon^2}{(1+\nu^*)^3(1-\epsilon)^2}}{\epsilon \frac{(1+\nu^*)^3-3(1+\nu^*)^2(\frac{1}{1-\epsilon}+\nu^*)}{(1+\nu^*)^6}} = \frac{(1+\nu^*)(1+\nu^*(1-\epsilon)^2)}{\epsilon(2(1+\nu^*(1-\epsilon))+\epsilon)(1-\epsilon)}$, which is positive for $0 \leq \underline{\epsilon} < \epsilon < \tilde{\epsilon} \leq 1$. Further, the implicit function theorem gives $\nu^{*'}(a) = -\frac{\frac{\partial F}{\partial a}}{\frac{\partial F}{\partial \nu^*}} = -\frac{\frac{f}{2da^2}}{\epsilon \frac{(1+\nu^*)^3-3(1+\nu^*)^2(\frac{1}{1-\epsilon}+\nu^*)}{(1+\nu^*)^6}} = \frac{f(1+\nu^*)^4(1-\epsilon)}{\epsilon 2da^2(2+2\nu^*+\epsilon-2\nu^*\epsilon)}$, which is positive for $0 \leq \underline{\epsilon} < \epsilon < \tilde{\epsilon} \leq 1$. Correspondingly, $\nu^{*'}(d) > 0$ and $\nu^{*'}(f) < 0$.

$N^* = \frac{1+\nu^*(1-\epsilon)}{1+\nu^*}a$ decreases in ϵ twofold: Directly because of $(1-\epsilon)$ and indirectly through the negative reaction on the increased ν^* (remember $\frac{\partial N(\nu, a, \epsilon)}{\partial \nu} < 0$ for $\epsilon > 0$ from footnote 29 in Chapter 2). Further, $\frac{dN^*}{da} = \frac{1+\nu^*(1-\epsilon)}{1+\nu^*} - \frac{a\epsilon\nu^{*'}(a)}{(1+\nu^*)^2}$. $\frac{dN^*}{da} > 0 \Leftrightarrow \frac{1+\nu^*(1-\epsilon)}{1+\nu^*} > \frac{a\epsilon}{(1+\nu^*)^2} \frac{f(1+\nu^*)^4(1-\epsilon)}{\epsilon 2da^2(2+2\nu^*+\epsilon-2\nu^*\epsilon)}$ where $\nu^{*'}(a)$ has been used from above. By using that $\frac{f}{2da} = \epsilon \frac{\frac{1}{1-\epsilon}+\nu^*}{(1+\nu^*)^3}$ from the implicit function for ν^* in (2.22), this simplifies to $(1+\nu^*(1-\epsilon)) > 0$, which is fulfilled for any $0 \leq \underline{\epsilon} < \epsilon < \tilde{\epsilon} \leq 1$. Thus, the direct positive effect of a on N^* outweighs the indirect negative effect of a through ν^* on N^* . For $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$, N^* increases in f and decreases in d through the negative reaction on ν^* . Thus, we have the comparative statics in (2.23).

B.1.3.4 Derivation of comparative statics on M^*

$M^* = (1+\nu^*)N^* = (1+\nu^*(1-\epsilon))a$ and thus $\frac{\partial M^*}{\partial \epsilon} > 0 \Leftrightarrow \frac{\partial(\nu^*(1-\epsilon))}{\partial \epsilon} > 0 \Leftrightarrow \frac{\partial \nu^*}{\partial \epsilon} \frac{\epsilon}{\nu^*} \geq \frac{\epsilon}{1-\epsilon}$ where the left hand side is the elasticity of ν^* with respect to ϵ . Using $\frac{\partial \nu^*}{\partial \epsilon}$ from the implicit function theorem above and simplifying reduces the inequality to $\nu^*(3\epsilon-1) < 1$. This corresponds then to $\nu^* > \frac{1}{3\epsilon-1}$ for $\epsilon < \frac{1}{3}$ (fulfilled since $\nu^* > 0$) and to $\nu^* < \frac{1}{3\epsilon-1}$ for $\epsilon > \frac{1}{3}$ (not necessarily fulfilled). Furthermore, $M^* = (1+\nu^*(1-\epsilon))a$ indicates that M^* is twofold increasing in a : Due to the direct effect of a and an indirectly through ν^* . Finally, M^* increases in d and decreases in f like ν^* does for $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$. Thus, we have the comparative statics in (2.24).

B.1.3.5 Derivation of comparative statics on the cost function

For given a and $\epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$, $\frac{dC_{tot}(\nu^*, \epsilon)}{d\epsilon} = 2a^2d \frac{(1+\nu^*(1-\epsilon))(-(\nu^*+\nu^{*2}+\epsilon\nu^{*'}(\epsilon)))}{(1+\nu^*)^3} + af(\nu^{*'}(\epsilon)(1-\epsilon) - \nu^*)$. By using equation (2.22) and simplifying, it follows that $\frac{dC_{tot}(\nu^*, \epsilon)}{d\epsilon} < 0 \Leftrightarrow \nu^*(1+\nu^*(1-\epsilon)) > 0$ which holds $\forall \nu^* > 0$ and $\forall \epsilon \in (\underline{\epsilon}, \tilde{\epsilon})$. In addition, for $\epsilon \geq \tilde{\epsilon}$, $\frac{dC_{tot}(N^*, \nu^*, \epsilon)}{d\epsilon} < 0$ because $\frac{d\nu^*}{d\epsilon} < 0$, but for $\epsilon \leq \underline{\epsilon}$, $\frac{dC_{tot}(N^*, \nu^*, \epsilon)}{d\epsilon} = 0$ because $\nu^* = 0$. Further, $\frac{dC_{tot}(N^*, \nu^*, \epsilon)}{da} = \frac{d\alpha}{da}wI^*\tilde{c} + 2d\frac{dN^*}{da}N^* + f\frac{dM^*}{da} > 0$ because $\frac{d\alpha}{da} > 0$, $\frac{dN^*}{da} \geq 0$ and $\frac{dM^*}{da} > 0$.

B.2 Comparative statics of fundamentals

Table B.1: Comparative statics of fundamentals

Parameter	I^*	I	w	γ	\tilde{R}	σ^2	\tilde{c}	c^s	f	d
on τ^*	+	.	+	-	+	-	-	+	-	-
on α^*	+	.	+	-	+	-	-	.	-	-
on K_s^*	-	+	?	+	-	+	+	.	+	+
on K_j^*	.	+	+	-	+	-

Notes: + is a positive and - a negative, ceteris paribus, comparative-static effect. ? indicates an ambiguous effect, which depends on the parameter values and . indicates no effect. The comparative-static effects of γ , \tilde{R} and σ^2 on τ^* are derived by using that the parameters satisfy $2f\gamma\sigma^2 < I^*\tilde{R}(\tilde{R} - 2\tilde{c})w$ (see Proposition 2.1). K_j^* for all $j = 1, \dots, N^*$.

B.3 Equilibria equations

B.3.1 Equilibria with correlated derivatives and neglected correlation

Table B.2: Equilibrium characterizations with derivatives and neglected correlation

	D	D,NC
τ^* :	$\tilde{c}\alpha^*(N_D^*, \nu) + c^s + \frac{dN_D^{*2} + f(1+\nu)N_D^*}{I^*w}$	$\tilde{c}\alpha^*(N_{D,NC}^*, \nu, \epsilon) + c^s + \frac{dN_{D,NC}^{*2} + f(1+\nu)N_{D,NC}^*}{I^*w}$
α^* :	$N_D^* \frac{\tilde{R}}{\gamma\sigma^2}$	$\frac{(1+\nu)N_{D,NC}^*}{1+\nu(1-\epsilon)} \frac{\tilde{R}}{\gamma\sigma^2}$
K_s^* :	$(1 - N_D^* \frac{\tilde{R}}{\gamma\sigma^2})wI$	$(1 - \frac{(1+\nu)N_{D,NC}^*}{1+\nu(1-\epsilon)} \frac{\tilde{R}}{\gamma\sigma^2})wI$
K_j^* :	$\frac{\tilde{R}}{\gamma\sigma^2}wI$	$\frac{1+\nu}{1+\nu(1-\epsilon)} \frac{\tilde{R}}{\gamma\sigma^2}wI$

Notes: K_j^* for all $j = 1, \dots, N^*$.

B.3.2 Planned and true utility levels

Utility level in benchmark model:

$$v^*(N^*, \tau^*) = \alpha^*(N^*)\tilde{R} + r - \frac{\gamma\alpha^*(N^*)^2\frac{\sigma^2}{N^*}}{2} - \tau^*$$

Utility level in model with derivatives:

$$v^*(N_D^*, \nu, \tau_D^*) = \alpha^*(N_D^*)\tilde{R} + r - \frac{\gamma\alpha^*(N_D^*)^2\frac{\sigma^2}{N_D^*}}{2} - \tau_D^*$$

Planned utility level in model with derivatives and neglected correlation:

$$v^*(N_{D,NC}^*, \nu, \epsilon, \tau_{D,NC}^*)^p = \alpha^*(N_{D,NC}^*, \nu, \epsilon)\tilde{R} + r - \frac{\gamma\alpha^*(N_{D,NC}^*, \nu, \epsilon)^2\frac{1+\nu(1-\epsilon)\sigma^2}{(1+\nu)N_{D,NC}^*}}{2} - \tau_{D,NC}^*$$

True utility level in model with derivatives and neglected correlation:

$$v^*(N_{D,NC}^*, \nu, \epsilon, \tau_{D,NC}^*)^t = \alpha^*(N_{D,NC}^*, \nu, \epsilon)\tilde{R} + r - \frac{\gamma\alpha^*(N_{D,NC}^*, \nu, \epsilon)^2\frac{\sigma^2}{N_{D,NC}^*}}{2} - \tau_{D,NC}^*$$

B.4 Simulation for endogenizing ν

B.4.1 Simulation procedure

This appendix describes the numerical algorithm to simulate the general equilibrium with endogenized ν :

- Set numerical values for the exogenous parameters $\mathbb{E}(R)$, r , σ^2 , I , w , γ , I^* , c^s , c^r , d and f .
- For $\epsilon = 0$ set $a_{\min}^* = N^*$ where N^* is the equilibrium level of diversification from the baseline models given in equation (2.12). And for $\epsilon = 1$ calculate a_{\max}^* where a_{\max}^* is $\arg \max_a \alpha_{\max} \tilde{R} + r - \frac{\gamma\alpha_{\max}^2\sigma^2}{2a} - \frac{\alpha_{\max}\tilde{c}wI^* + c^s w I^* + d + fa}{wI^*}$ where $\alpha_{\max} = \max\{a\frac{\tilde{R}}{\gamma\sigma^2}, 1\}$. $N_{\min} = 1$ and $\nu_{\epsilon=1} = a - 1$ was used.
- Set the step sizes s_ϵ for $\epsilon \in [0, 1]$ and s_a for the perceived level of diversification $a \in [a_{\min}^*, a_{\max}^*]$.¹
- Solve $\max_{\alpha \in [0, 1]} \alpha \mathbb{E}(R) + (1 - \alpha)r - \frac{\gamma\alpha^2\sigma^2}{2a} - \tau$ (where $\sigma_B^2 = \frac{\sigma^2}{a}$ was used) to get the risky portfolio share as a function of a : $\tilde{\alpha}^*(a) = a\frac{\tilde{R}}{\gamma\sigma^2}$.

¹Since C_{tot} is monotone in ϵ and a (see Appendix B.1.3.5) the simulation results are monotone in ϵ and a .

- e) For each ϵ run the following steps:
- (a) For each a run the following steps:
 - i. Re-set $\alpha^*(a) = \max\{\tilde{\alpha}^*(a), 1\}$.
 - ii. Minimize $C_{tot}(N, \nu, \epsilon) = c^r \alpha^*(a) w I^* + c^s (1 - \alpha^*(a)) w I^* + d N^2 + f(1 + \nu) N$ with respect to ν and N subject to $a = \frac{(1+\nu)N}{1+\nu(1-\epsilon)}$, $N \geq 1$ and $\nu \geq 0$.
 - iii. Store the solution (N^*, ν^*) from 5.(a).ii. In addition store the minimal costs $C_{tot}(N^*, \nu^*, \epsilon)$ and $\tau^* = \frac{C_{tot}(N^*, \nu^*, \epsilon)}{I^* w}$.
 - iv. Calculate households utility $v^*(a) = \alpha^*(a) \mathbb{E}(R) + (1 - \alpha^*(a)) r - \frac{\gamma \alpha^{*2} \sigma^2}{2a} - \tau^*$ and store it.
 - (b) Determine the utility maximizing a^* by picking the maximum value $v^*(a^*)$ of the household's utility level stored in 5.(a).iv.
 - (c) Pick the to a^* corresponding values ν^* , N^* and τ^* from 5.(a).iii. Derive $M^* = (1 + \nu^*) N^*$. Calculate the true and the planned bundle variance (i.e., $\sigma_B^{*2} = \frac{\sigma^2}{N^*}$ and $\sigma_B^{*2}(N^*, \nu^*, \epsilon) = \frac{\sigma^2}{a^*}$, respectively).
 - f) Plot the values of a^* , ν^* , N^* , M^* , σ_B^{*2} , $\sigma_B^{*2}(N^*, \nu^*, \epsilon)$ and τ^* in dependence of $\epsilon \in [0, 1]$.

B.4.2 Simulation results: Equilibrium effects of ϵ

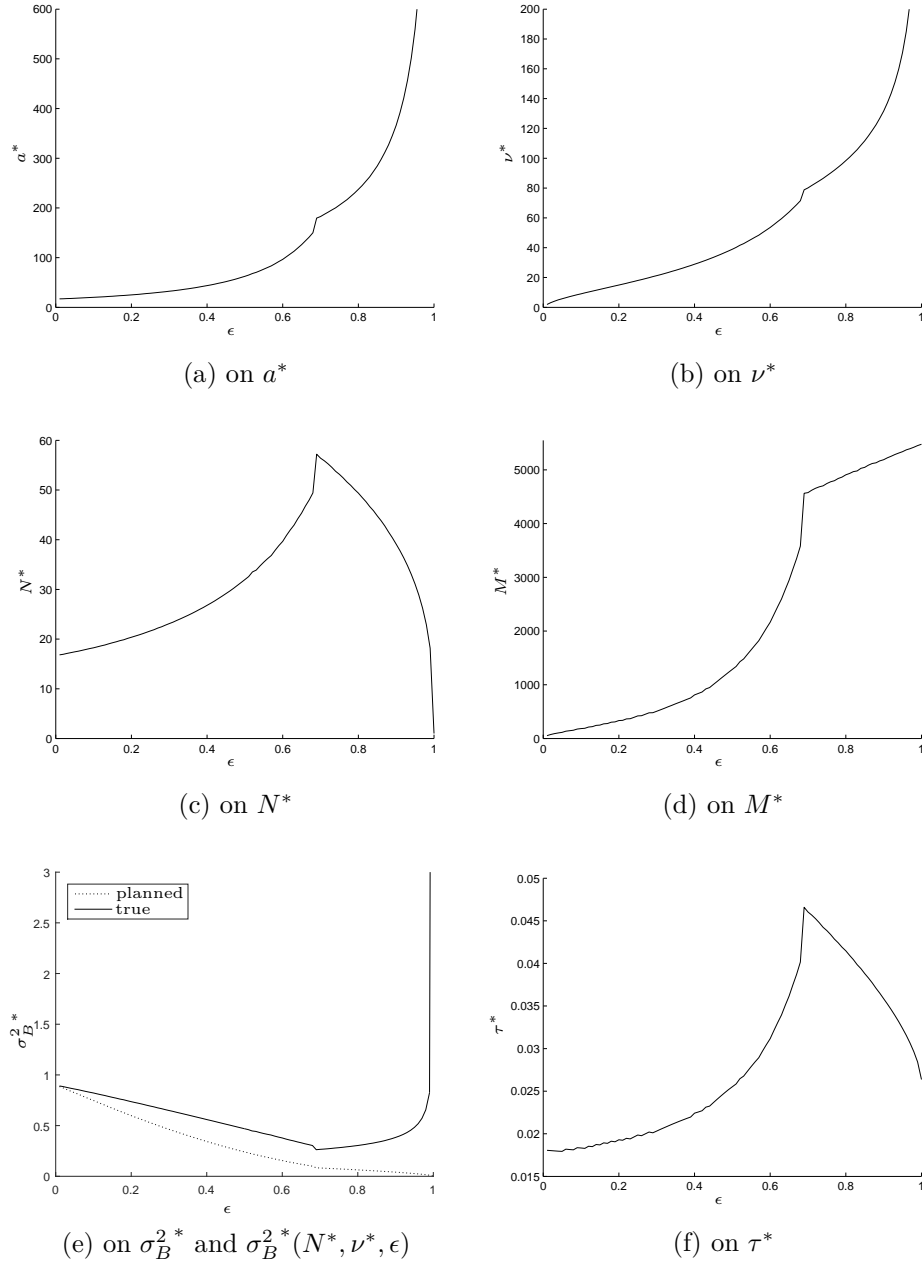


Figure B.2: General equilibria for $\epsilon \in [0, 1]$

Notes: Simulation results for parameter values: $r = 1.01$ (i.e., approximately the fed funds rate in 2010), $\mathbb{E}(R) = 1.1$ reflects a 10% risky interest rate, $w = 200,000$ (i.e., approximately the total net worth of \$60,000bn in 2010 divided by number of citizens 300m in the U.S.), $\gamma = 1$ (standard risk aversion parameter), $\sigma^2 = 15$ (i.e., approximately the VIX volatility index in 2010), $c^s = 0.015$ and $c^r = .025$ (weighted sum corresponds approximately to 2% cost of financial intermediation as estimated in Philippon (2015)), $I = 60,000,000$ (i.e., approximately the 20% share, who invest in stocks, of all citizens 300m in the U.S.), $d = 5,000$ and $f = 200$ (i.e., approximately a month and a day salary, respectively, of a financier in the U.S.), $n = 15,000$ (i.e., approximately the number of banks in the U.S. in 2010). $\alpha \in [0, 1]$ was considered. Note that the kink in panel (c) is because for larger ϵ the positive a -effect on N will be outweighed by the negative ϵ -effect on N . For the same reasons there are kinks in the other panels.

C Appendix: Chapter 3

C.1 Timing

The timing of the activities within a period T is as follows (see Figure C.1). At first, a mass of $1 - \Delta$ households is born. An endogenously determined share λ of them decides to become an entrepreneur and the rest decides to be a worker. The newborn entrepreneurs sign a lifetime binding financial contract with the banks. The banks give loans $b(V^E)$, according to the amount entitled by the respective terms of the contract, to all entrepreneurs in the economy. Entrepreneurs pay the production costs of capital and labor input with the bank loans and production takes place under uncertainty. (Workers consume and save by buying annuities from the banks from their labor income and capital returns). After production, entrepreneurs observe the state of their productivity realization and make a report about it to the bank. Then, entrepreneurs make state-contingent repayments $m_s(V^E)$ to the banks according to their financial contract and consume the remaining net production revenue. Further, the contract determines state-contingent promised values $V_s^E(V^E)$ as future state variable. Finally, a share $1 - \Delta$ of the workers and entrepreneurs dies and the associated firms exit.

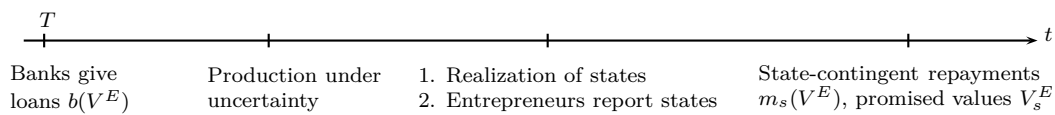


Figure C.1: Timing of terms of financial contract within one period

C.2 Derivations

C.2.1 Derivations of financial contract properties

The proofs for Proposition 3.1 and 3.2 and Lemma 3.1 and 3.2 follow the proofs on the optimal social insurance in Ljungqvist and Sargent (2000) which are based on Thomas and Worrall (1990).

C.2.1.1 Proof of Proposition 3.1

Proof. Using the definition in (3.12) and summing up $C_{s,s-1} + C_{s-1,s}$ we conclude from (IC): $C_{s,s-1} + C_{s-1,s} \geq 0$, which is equivalent to

$$\begin{aligned} U(\theta_s R(b) - m_s, L^E) - U(\theta_{s-1} R(b) - m_{s-1}, L^E) \geq \\ U(\theta_{s-1} R(b) - m_s, L^E) - U(\theta_{s-1} R(b) - m_{s-1}, L^E) \end{aligned} \quad (\text{C.1})$$

Since $\theta_s > \theta_{s-1}$ and given the strict concavity of the utility function in consumption, (C.1) is satisfied only if $m_s \geq m_{s-1}$. It then follows from $C_{s,s-1} \geq 0$ that $V_s^E \geq V_{s-1}^E$. \square

C.2.1.2 Proof of Lemma 3.1

Proof. Without loss of generality, we prove from the local downward constraints $C_{s,s-1} \geq 0, \forall s \in \mathcal{S}$, that for any $i > j, i, j \in \mathcal{S}, C_{i,j} \geq 0$. The case of $i < j$ can be proved from the local upward constraints $C_{s,s+1} \geq 0, \forall s \in \mathcal{S}$, using the same logic.

Proof with mathematical induction: For $n = 1, C_{j+n,j} \geq 0$ holds according to the local downward constraint. Suppose for $n \geq 1, C_{j+n,j} \geq 0, \forall j \in \mathcal{S}$ holds; we need to prove that $C_{j+n+1,j} \geq 0$. For simplicity of notation denote $i = j + n$.

First, $C_{i,j} \geq 0$ and $C_{i+1,i} \geq 0$ are equivalent to the following inequalities:

$$\begin{aligned} U(\theta_i R(b) - m_i, L^E) + \beta \Delta V_i^E - U(\theta_i R(b) - m_j, L^E) - \beta \Delta V_j^E \geq 0, \\ U(\theta_{i+1} R(b) - m_{i+1}, L^E) + \beta \Delta V_{i+1}^E - U(\theta_{i+1} R(b) - m_i, L^E) - \beta \Delta V_i^E \geq 0. \end{aligned}$$

Summing up the two inequalities we have:

$$\begin{aligned} U(\theta_{i+1} R(b) - m_{i+1}, L^E) + \beta \Delta V_{i+1}^E - \beta \Delta V_j^E + \\ U(\theta_i R(b) - m_i, L^E) - U(\theta_i R(b) - m_j, L^E) - U(\theta_{i+1} R(b) - m_i, L^E) \geq 0. \end{aligned} \quad (\text{C.2})$$

Using the strict concavity of the utility function, the fact $\theta_{i+1} > \theta_i$, and $m_i \geq m_j$ from Proposition 3.1, we have additionally the following inequality:

$$\begin{aligned} U(\theta_{i+1} R(b) - m_i, L^E) - U(\theta_{i+1} R(b) - m_j, L^E) \geq \\ U(\theta_i R(b) - m_i, L^E) - U(\theta_i R(b) - m_j, L^E) \end{aligned} \quad (\text{C.3})$$

Adding (C.3) to (C.2) we have

$$U(\theta_{i+1} R(b) - m_{i+1}, L^E) + \beta \Delta V_{i+1}^E - \beta \Delta V_j^E - U(\theta_{i+1} R(b) - m_j, L^E) \geq 0.$$

Namely, $C_{i+1,j} \geq 0$. \square

C.2.1.3 Proof of Lemma 3.2

Proof. First, we prove by contradiction that the local downward constraints must bind: Suppose that there exists an optimal contract $\{b, m_s, V_s^E\}_{s \in \mathcal{S}}$ such that for some $i \in \mathcal{S}$ the downward constraint does not bind (i.e., $C_{i,i-1} > 0$). Then, the general procedure is as follows: We prove that there exists a mean-preserving contraction transformation on $\{V_j^E\}_{j=i, \dots, S}$ such that the new contract $\{b, m_s, \hat{V}_s^E\}_{s \in \mathcal{S}}$, where $\hat{V}_j^E = V_j^E$, for $j = 1, 2, \dots, i-1$, fulfills all constraints. In particular, we make a transformation with $\sum_{s \in \mathcal{S}} \pi_s \hat{V}_s^E = \sum_{s \in \mathcal{S}} \pi_s V_s^E$, and $\hat{V}_j^E - \hat{V}_l^E \leq V_j^E - V_l^E$, $\forall j, l \in \mathcal{S}$, with at least one pair of $\{j, l\}$ giving strict inequality. In this case, under the assumption that $P(V^E)$ is strictly concave, the banks' profit increases strictly with the new contract. This contradicts the fact that $\{b, m_s, V_s^E\}_{s \in \mathcal{S}}$ is an optimal contract.

Now we describe explicitly the procedure of performing a mean-preserving contraction transformation on the contract:

Keeping $\{m_{i-1}, m_i, V_{i-1}^E\}$ as before, we decrease V_i^E until $C_{i,i-1} = 0$. Since changing V_i^E will influence the local downward incentive constraints for $s = i+1$ and sequentially $s = i+2, \dots, S$, we decrease for each $s = i+1, \dots, S$, V_s^E such that $C_{s,s-1} = 0$. As a result we have a new sequence of future promised value $\{V_s^{E'}\}_{s \in \mathcal{S}} = \{V_1^E, V_2^E, \dots, V_{i-1}^E, V_i^{E'}, V_{i+1}^{E'}, \dots, V_S^{E'}\}$. Now we add a positive constant, \bar{v} , to the sequence of future promised value, such that the promise keeping constraint is regained. Let $\hat{V}_s^E = V_s^{E'} + \bar{v}$. We have a new contract $\{b, m_s, \hat{V}_s^E\}_{s \in \mathcal{S}}$.

First, note that the new contract fulfills the local upward constraints automatically given the strict concavity of the utility function and the fact that $C_{s,s-1} = 0 \forall s \in \mathcal{S}$ (see argumentation in the last part of this proof). In addition, the promise keeping constraint is still fulfilled due to the mean-preserving transformation, and the limited liability constraints are uninfluenced since b and $\{m_s\}_{s \in \mathcal{S}}$ are unchanged. Finally, for any $j = i, \dots, S$, V_{j+1}^E must decrease at least as much as V_j^E to guarantee that $C_{j+1,j} = 0$. Therefore, for any $j = i, \dots, S$, $\bar{v} \leq V_j^E - V_j^{E'}$, indicating that $\hat{V}_j^E \leq V_j^E$ and remember that for $j = 1, 2, \dots, i-1$ $\hat{V}_j^E = V_j^E$. Since $\{V_s^E\}_{s \in \mathcal{S}}$ fulfills the credibility constraints, so does the new contract.

Further, notice from the procedure that the gap of the promised values between two successive states, s and $s-1$, is either unchanged or decreased, with a definite decrease in $\hat{V}_i^E - \hat{V}_{i-1}^E$. Following this we know $\forall j, l \in \mathcal{S}$, $V_j^E - V_l^E$ is non-increasing. Thus, the new contract is a mean-preserving contraction. This contradicts that $\{b, m_s, V_s^E\}_{s \in \mathcal{S}}$ is an optimal contract: We know that the local downward constraints always bind.

Given that $C_{s,s-1} = 0, \forall s \in \mathcal{S}$, rewriting the constraint we have

$$\beta \Delta(V_s^E - V_{s-1}^E) = U(\theta_s R(b) - m_{s-1}, L^E) - U(\theta_s R(b) - m_s, L^E)$$

Since $\theta_{s-1} < \theta_s, m_{s-1} \leq m_s$ and the utility function is strictly concave, we have

$$\begin{aligned} U(\theta_{s-1} R(b) - m_{s-1}, L^E) - U(\theta_{s-1} R(b) - m_s, L^E) &\geq \\ U(\theta_s R(b) - m_{s-1}, L^E) - U(\theta_s R(b) - m_s, L^E) &= \beta \Delta(V_s^E - V_{s-1}^E), \end{aligned}$$

where strict inequality holds for $m_{s-1} < m_s$. Therefore, we have directly from this that the local upward constraint is never binding. Namely, $C_{s-1,s} > 0, \forall s \in \mathcal{S}$. \square

C.2.1.4 Proof of Proposition 3.2

Proof. The non-decreasing entrepreneurs' utility is direct result of the binding local downward constraints.

The non-decreasing profit of banks is proved by contradiction. Suppose for the optimal contract, $\{b, m_s, V_s^E\}_{s \in \mathcal{S}}$, there exists $i, j \in \mathcal{S}, i > j$, such that

$$-b + m_i + \frac{\Delta}{1+r} P(V_i^E) < -b + m_j + \frac{\Delta}{1+r} P(V_j^E).$$

Substituting (m_i, V_i^E) with (m_j, V_j^E) increases banks' profit in state i . Since the downward constraint binds, $C_{i,j} = 0$, the terms of contract, (m_j, V_j^E) , entitle the entrepreneurs the same promised value as (m_i, V_i^E) . This means that we find an improvement that increases the profit of the banks without violating any constraints. This contradicts the optimality of the original contract. Therefore, in the optimal contract the banks' profits cannot decline with a higher productivity realization. \square

C.2.1.5 Proof of Proposition 3.3

Proof. We use the Lagrangian. For simplification, we derive the Lagrangian and the F.O.C. for the case of two states (i.e., $\mathcal{S} = \{l, h\}$ with $\pi_h = \pi, \pi_l = 1 - \pi$). The

Lagrangian is given by:

$$\begin{aligned}
\mathcal{L} = & \max_{\{b, m_s, V_s^E\}_{s \in S}} -b + \sum_{s=\{l, h\}} \pi_s \left[m_s + \frac{\Delta}{1+r} P(V_s^E) \right] \\
& + \lambda_1 \left\{ \sum_{s=\{l, h\}} \pi_s [U(c_{ss}, L^E) + \beta \Delta V_s^E] - V^E \right\} \\
& + \lambda_2 \left\{ U(c_{hh}, L^E) + \beta \Delta V_h^E - U(c_{hl}, L^E) - \beta \Delta V_l^E \right\} \\
& + \lambda_3 \left\{ U(c_{ll}, L^E) + \beta \Delta V_l^E - U(c_{lh}, L^E) - \beta \Delta V_h^E \right\} \\
& + \lambda_4 c_{hh} + \lambda_5 c_{ll} \\
& + \lambda_6 (V_{max}^E - V_l^E) + \lambda_7 (V_l^E - V_{min}^E) \\
& + \lambda_8 (V_{max}^E - V_h^E) + \lambda_9 (V_h^E - V_{min}^E)
\end{aligned}$$

where $c_{ij} = \theta_i R(b) - m_j$. $\lambda_i \geq 0$, $i = 1, \dots, 9$ are the Lagrangian multiplier for (PK), (IC), (LL) and (CC), respectively. The F.O.C.s are

$$\frac{\partial \mathcal{L}}{\partial m_h} = \pi - (\lambda_1 \pi + \lambda_2) U'(c_{hh}) + \lambda_3 U'(c_{lh}) - \lambda_4 = 0; \quad (C.4)$$

$$\frac{\partial \mathcal{L}}{\partial m_l} = (1 - \pi) - (\lambda_1(1 - \pi) + \lambda_3) U'(c_{ll}) + \lambda_2 U'(c_{hl}) - \lambda_5 = 0; \quad (C.5)$$

$$\frac{\partial \mathcal{L}}{\partial V_h^E} = \frac{\pi}{1+r} P'(V_h^E) + (\lambda_1 \pi + \lambda_2 - \lambda_3) \beta - \lambda_8 + \lambda_9 = 0; \quad (C.6)$$

$$\frac{\partial \mathcal{L}}{\partial V_l^E} = \frac{1-\pi}{1+r} P'(V_l^E) + (\lambda_1(1 - \pi) - \lambda_2 + \lambda_3) \beta - \lambda_6 + \lambda_7 = 0; \quad (C.7)$$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial b} = & -1 + \{(\lambda_1 \pi + \lambda_2) U'(c_{hh}) - \lambda_2 U'(c_{hl}) + \lambda_4\} \theta_h R'(b) \\
& + \{(\lambda_1(1 - \pi) + \lambda_3) U'(c_{ll}) - \lambda_3 U'(c_{lh}) + \lambda_5\} \theta_l R'(b) = 0.
\end{aligned} \quad (C.8)$$

These together with the complementary conditions and $\lambda_i \geq 0$, $i = 1, \dots, 9$ characterize the conditions an optimal contract needs to fulfill. Furthermore, we use that the downward constraint always binds ($\lambda_2 \geq 0$) whereas the upward constraint for $m_s > m_{s-1}$ never does ($\lambda_3 = 0$).

From (C.4), (C.5) and (C.8) follows

$$\mathbb{E}(\theta) R'(b) + [\lambda_3 U'(c_{lh}) - \lambda_2 U'(c_{hl})] (\theta_h - \theta_l) R'(b) = 1$$

Since $\lambda_3 = 0$ we have

$$\mathbb{E}(\theta) R'(b) + [-\lambda_2 U'(c_{hl})] (\theta_h - \theta_l) R'(b) = 1. \quad (C.9)$$

Since $\lambda_2 \geq 0$ we have that $\mathbb{E}(\theta)R'(b) \geq 1$ and thus $b \leq b^*$, where b^* is the efficient bank loan level defined in (3.15).

Suppose now that the credibility constraints (CC) are not binding so that we have $\lambda_i = 0$, $i = 6, \dots, 9$. Then, it follows from (C.6) and (C.7) that $\lambda_2 = \frac{\pi(1-\pi)}{(1+r)\beta}(P'(V_l^E) - P'(V_h^E))$, which is positive for $V_h^E > V_l^E$ because $P(V_h^E)$ is strictly concave. Thus, the level of $b(V^E)$ is the result of the functional forms of the utility, the production and the value function and their relative curvature compared to each other.

□

C.2.2 Derivation of the goods market clearing condition

The goods market is cleared if aggregate output equals the sum of consumption of all households and aggregates investment (i.e., replacement of depreciated capital in stationary case). Remember from (3.29) that the formal condition is $\lambda Y = \lambda C^E + (1 - \lambda)C^W + \lambda \delta K^D$. To prove that the goods market clearing condition can be derived from the other equations, we need only to prove that the RHS of (3.29) can be simplified to λY .

We aggregate the consumption of entrepreneurs C^E and workers C^W . Plugging in entrepreneur's consumption $c(\cdot)$ from (3.7) into (3.24) (using $R(b) = R(b(V^W))$ and $m_s = m(V^E, \theta_s)$) gives aggregate consumption of entrepreneurs:

$$C^E = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^{\tau} \sum_{s \in S} \pi_s \int [\theta_s R(b(V^E)) - m(V^E, \theta_s)] d\Psi_{\tau}(V^E) = Y - M, \quad (\text{C.10})$$

where the second equality follows from (3.23) and (3.22).

A cohort τ worker's consumption is $c_{\tau} = wl_{\tau} + A_{\tau} - p^A A_{\tau+1}$ (follows from the budget constraint (3.3)). Aggregation gives

$$\begin{aligned} C^W &= \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^{\tau} (wl_{\tau} + A_{\tau} - p^A A_{\tau+1}) \\ &= wL^S - D + \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^{\tau} A_{\tau} \\ &= wL^S - D + \Delta \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^{\tau-1} A_{\tau} \\ &= wL^S - D + \Delta \frac{1}{p^A} D \\ &= wL^S + rD, \end{aligned} \quad (\text{C.11})$$

where the second and forth equality follow from (3.18) and (3.17) and the annuity price $p^A = \frac{\Delta}{1+r}$ from (3.1) is used.

Entrepreneurs have the bank loans to finance the production costs. We assume that there is no moral hazard problem (i.e., the money is used optimally for production costs). Thus, the constraint in entrepreneurs' decision problem in (3.5) is binding so that we have $b(V^E) = wl^*(V^E) + (r + \delta)k^*(V^E)$. In aggregation this means

$$B = \sum_{\tau=0}^{\infty} (1 - \Delta) \Delta^{\tau} \int b(V^E) d\Psi_{\tau}(V^E) = wL^D + (r + \delta)K^D. \quad (\text{C.12})$$

Plugging (C.10) and (C.11) into the RHS of (3.29) gives:

$$\begin{aligned} \text{RHS} &= \lambda(Y - M) + (1 - \lambda)(wL^S + rD) + \lambda\delta K^D \\ &= \lambda Y - \lambda M + (1 - \lambda)wL^S + (1 - \lambda)rD + \lambda rE - \lambda rE + \lambda\delta K^D \\ &= \lambda Y + \lambda wL^D + \lambda(r + \delta)K^D - \lambda(M + rE) \\ &= \lambda Y + \lambda B - \lambda B \\ &= \lambda Y = \text{LHS}, \end{aligned}$$

where the equilibrium conditions (3.27) and (3.28) were used in the third equality and (3.25) and (C.12) were used in the third. This closes the proof that the goods market clearing condition can be derived from clearing in the capital and labor markets.

C.3 Numerical procedure for optimization problems

In this section, the numerical dynamic programming procedures for solving the optimization problems are described. Specifically, we describe the algorithm for workers' decision problem (Section C.3.1) and banks' optimal contract (Section C.3.2), for given r and w .

C.3.1 Numerical procedure: Workers

- a) Use constant r and w .
- b) Set a grid for the state variable A . A_{grid} denote these grid points of A . We set $A = [0, 10]$ and generate $nA = 50$ Chebyshev grid points on the interval.¹ We manually replace the lowest Chebyshev point with the lower bound $A = 0$.

¹See footnote 26 in Chapter 3 for an explanation of Chebyshev points.

- c) Give an initial guess for the functional form of the value function, $V^W(A; r, w)^0$, of the policy functions $l(A)^0$ and $A'(A)^0$ and of $c(A)^0$. Notice that in the numerical exercise functions are defined on the discrete grid points *Agrid*. Namely, it is a mapping of each grid point into a number. We use $V^W(A_i)^0 = -\exp(-A_i) - 0.1$, $c(A_i)^0 = 0.1$, $l(A_i)^0 = 0.6$ and $A'(A_i)^0 = 0$ for each $A_i \in Agrid$, $i \in \{1, \dots, nA\}$.
- d) Solve on each grid point $A_i \in Agrid$, $i \in \{1, \dots, nA\}$, the worker's problem in (3.2) subject to (3.3), $c(A_i) \geq 0$, $l(A_i) \in [0, 1]$ and $A'(A_i) \geq 0$. This gives us the optimal solution of the system, $\{c(A_i)^1, l(A_i)^1, A'(A_i)^1\}$ and the corresponding updated value function $V^W(A_i; r, w)^1$ at $A_i \in Agrid$, $i \in \{1, \dots, nA\}$. We apply the *fmincon*-command in Matlab, which finds the minimum of a constrained nonlinear multivariable function using the interior point algorithm. To calculate the updated value function, we use spline interpolation (i.e., cubic interpolation of the values of neighbor-points) on $V^W(A; r, w)^0$ to get values for $A'(A)$ which lie between two *Agrid*-points.
- e) Compare the two successive iterations of value functions, $V^W(A; r, w)^1$ with $V^W(A; r, w)^0$, by defining a distance measure d_{VW} . If $d_{VW} \leq \epsilon_{VW}$, then take the current iteration of the value function and policy functions as the solution and go to f). If $d_{VW} > \epsilon_{VW}$ start over with step c) by updating $V^W(A; r, w)^0 = V^W(A; r, w)^1$ and $c(A)^0 = c(A)^1$, $l(A)^0 = l(A)^1$ and $A'(A)^0 = A'(A)^1$ as the new starting values for the next iteration. We use $d_{VW} \equiv \max_{i \in \{1, \dots, nA\}} |V^W(A_i; r, w)^1 - V^W(A_i; r, w)^0|$. We set as criterion for ending the iterations the tolerated distance $\epsilon_{VW} = 0.0001$.
- f) Save the value function $V^W(A; r, w) = V^W(A; r, w)^1$, the policy functions $A'(A) = A'(A)^1$ and $l(A) = l(A)^1$ and $c(A) = c(A)^1$.

C.3.2 Numerical procedure: Financial contract

- a) Use constant r and w .
- b) Set a grid for the state variable V^E on the interval $[V_{min}^E, V_{max}^E]$. V^E_{grid} denote these grid points of V^E . We generate $nV^E = 50$ Chebyshev grid points on the interval.² We manually replace the lowest Chebyshev point with the lower bound V_{min}^E .

²See footnote 26 for an explanation of Chebyshev points.

- c) Make an initial guess of the functional form of the value function, $P(V^E; r, w)^0$, and of the policy functions, $\{b(V^E)^0, m_s(V^E)^0, V_s^E(V^E)^0\}_{s \in \{h, l\}}$. Notice that in the numerical exercise functions are defined on the discrete grid points $V^E grid$. Namely, it is a mapping of each grid point into a number. We use $P(V_i^E; r, w)^0 = \log(-V_i^E)$, $b(V_i^E)^0 = 1$, $m_h(V_i^E)^0 = 3$, $m_l(V_i^E)^0 = 1$, $V_h^E(V_i^E)^0 = V_l^E(V_i^E)^0 = V_i^E$ for each $V_i^E \in V^E grid$, $i \in \{1, \dots, nV^E\}$.
- d) Solve for each $V_i^E \in V^E grid$, $i \in \{1, \dots, nV^E\}$ the optimal contract in (3.11) subject to (PK), (IC), (LL) and (CC).³ This gives the optimal contract at each $V_i^E \in V^E grid$, $i \in \{1, \dots, nV^E\}$, $\{b(V_i^E)^1, m_s(V_i^E)^1, V_s^E(V_i^E)^1\}_{s \in \{h, l\}}$, and the corresponding updated value function, $P(V_i^E; r, w)^1$. We apply the *fmincon*-command in Matlab, which finds the minimum of a constrained nonlinear multivariable function using the interior point algorithm. To calculate the value function for the next iteration, we use spline interpolation (i.e., cubic interpolation of the values of neighbor-points) on $P(V_i^E; r, w)^0$ to get values for $V_h^E(V^E)^1$ and $V_l^E(V^E)^1$ which lie between two grid points.
- e) Compare the two successive iterations of value functions, $P(V_i^E; r, w)^1$ with $P(V_i^E; r, w)^0$, by defining a distance measure d_P . If $d_P \leq \epsilon_P$, then take the current iteration of the value function and policy functions as the solution and go to f). If $d_P > \epsilon_P$, start over with step c) by updating $P(V^E; r, w)^0 = P(V^E; r, w)^1$ and $b(V^E)^0 = b(V^E)^1$, $m_h(V^E)^0 = m_h(V^E)^1$, $m_l(V^E)^0 = m_l(V^E)^1$, $V_h^E(V^E)^0 = V_h^E(V^E)^1$ and $V_l^E(V^E)^0 = V_l^E(V^E)^1$ as the new starting value for the next iteration. We use $d_P \equiv \max_{i \in \{1, \dots, nV^E\}} |P(V_i^E; r, w)^1 - P(V_i^E; r, w)^0|$ and set as criterion for ending the iterations the tolerated distance $\epsilon_P = 0.0001$.
- f) Save the value function $P(V^E; r, w) = P(V^E; r, w)^1$ and the optimal contract $b(V^E) = b(V^E)^1$, $m_h(V^E) = m_h(V^E)^1$, $m_l(V^E) = m_l(V^E)^1$, $V_h^E(V^E) = V_h^E(V^E)^1$ and $V_l^E(V^E) = V_l^E(V^E)^1$.

C.4 Simulation procedure for general equilibrium

In this appendix, we describe how the two partial parts from above and the entrepreneurs' decision are combined to calculate the general equilibrium results.

- a) We simulate for $N^E = 10,000,000$ entrepreneurs' life paths with history of productivity realizations and age (i.e., realization of productivity state and death /

³For (IC) we put in the constraint only the binding local downward constraint, since by the result of Lemma 3.2 the local upward constraint is never binding for the optimal contract.

survival): For each entrepreneur, while surviving, we draw on the interval between 0 and 1 (uniformly distributed) randomly two separate numbers to simulate productivity and life shocks, u_θ and u_Δ , respectively: If $u_\theta < \pi_l$ the entrepreneur faces a low productivity realization θ_l and otherwise a high productivity realization θ_h . If $u_\Delta > 1 - \Delta$ the entrepreneur survives, becomes one year older and continues production; if $u_\Delta \leq 1 - \Delta$ the entrepreneur dies and the firm exits the market.

- b) Give a grid on the range of interest rate and wage and start from a grid point (r, w) .
- c) Solve the optimization problems for this (r, w) : Workers' decision problem as described in C.3.1 and banks' optimal contract as described in C.3.2.
- d) Calculate the aggregate deposits D and the labor supply L^S of workers. More specifically, sum up the weighted savings and labor supply, $p_A A'(A_\tau)$ and $l(A_\tau)$ from c), according to (3.17) and (3.18) for ages $\tau = 0, \dots, N^W$. N^W is determined such that $\sum_{\tau=0}^{N^W} (1 - \Delta)\Delta^\tau \geq 1 - \epsilon_L$ with $\epsilon_L = 0.0001$. This indicates that the cohorts we aggregate cover approximately all workers in the economy.
- e) Set $V_0^E = V^W(0; r, w)$ from c). Using the simulation of the life paths from a), determine the corresponding promised value V^E for each entrepreneur using the transition function $V_s^E(V^E)$ from c) over its life path. Banks loans and repayments are determined by $\{b(V^E), m_s(V^E)\}_{s \in \{h, l\}}$ from c). Aggregations over all entrepreneurs are then the sums B and M of all banks loans b and repayments m , respectively, divided by N^E to normalize the mass of the population to 1. From B follows aggregate capital demand K^D and aggregate labor demand L^D directly according to (3.33).
- f) Determine the share of entrepreneurs from the labor market condition, $\lambda = \frac{L^S}{L^D + L^S}$.
- g) Check if the equilibrium conditions (zero-profit and capital market clearing) are close to zero, specifically: $|P(V_0^E; r, w)| \leq \epsilon_{eq,p}$ and $|\lambda K^D - (1 - \lambda)D - \lambda \frac{B-M}{r}| \leq \epsilon_{eq,c}$? If yes, go to h). If not, restart from c) with a new grid point (r', w') . When updating (r', w') , to get the direction to the new grid point, we use the observable fact that the banks' profit and the excess capital demand are both decreasing in r and w and that their gap increases in w and decreases in r (see Appendix C.5 for a detailed description of this algorithm). We set $\epsilon_{eq,p} = 0.0037$ and $\epsilon_{eq,c} = 0.0037$.
- h) Save the results.

C.5 Algorithm to find general equilibrium

When grid-searching for the equilibrium (r, w) , we observed (at least locally around the equilibrium values) that the banks' profit $\Pi \equiv P(V^W(0; r, w))$ and the excess capital demand $X \equiv \lambda K^D(r, w) - (1 - \lambda)D(r, w) - \lambda E(r, w)$ are both decreasing in r and w . This means their partial derivatives have a negative sign:

$$\Pi_r < 0, \quad \Pi_w < 0, \quad X_r < 0 \text{ and } X_w < 0 \quad (\text{C.13})$$

In a (r, w) -diagram, (C.13) implies two properties of the iso-profit and iso-excess demand curves: (i) Both loci are downward sloping. (ii) A northeast shift of a locus decreases the corresponding value of the respective iso-curve. The slope of the iso-profit curve and of the iso-excess demand curve are given by $S_\Pi = -\frac{\Pi_w}{\Pi_r}$ and $S_X = -\frac{X_w}{X_r}$, respectively.

In addition to (C.13), we observe that the gap $G \equiv X - \Pi$ is decreasing in r and increasing in w . From the partial derivatives $G_r < 0$ and $G_w > 0$ we know that

$$\Pi_r > X_r \text{ and } X_w > \Pi_w. \quad (\text{C.14})$$

From (C.14) follows for the slopes of the iso-profit and iso-excess demand curves

$$|S_X| < |S_\Pi|. \quad (\text{C.15})$$

Thus, the iso-profit curve is steeper than iso-excess demand curve. This indicates single-crossing property of the two curves.

Using (C.13) and (C.15), we know that (at least locally) there exists a unique equilibrium. Furthermore, the properties of the iso-curves indicate the direction for approaching the equilibrium from any off-equilibrium point.

Figure C.2 exhibits an illustration of iso-profit and iso-excess demand curves and the algorithm to find the equilibrium. Suppose that for an initial guess of (r, w) the value of the iso-profit is $\bar{\pi}_A > 0$ (a point like A). Then, first, according to (C.13), we increase r (or w) until the zero-profit condition holds (i.e., on the locus of $\Pi = 0$). (Note that, for a point like B with $\bar{\pi}_B < 0$ the opposite argument than for a point like A applies). This first step could bring us to: (i) a point with negative excess capital demand like C ; (ii) a point with positive excess capital demand like D ; or (iii) a point with zero excess demand indicating the equilibrium $Eq.$. Second, at a point C with $X = \bar{x}_C < 0$, (C.13) and (C.15) suggest a south-east shift of (r, w) (i.e., $r \downarrow, w \uparrow$) along the locus of the iso-profit curve until the excess demand increases to 0 ($C \rightarrow Eq.$). Similarly, at a point D with $X = \bar{x}_D > 0$, a north-west shift of (r, w) (i.e., $r \uparrow, w \downarrow$) decreases the excess demand

and thus approaches the equilibrium ($D \rightarrow Eq.$). This algorithm finds the unique general equilibrium. In practice, we need to guess the change of (r, w) in each step. Depending on the accuracy of the guesses, more or less “back and forths” may be needed until the equilibrium is reached.

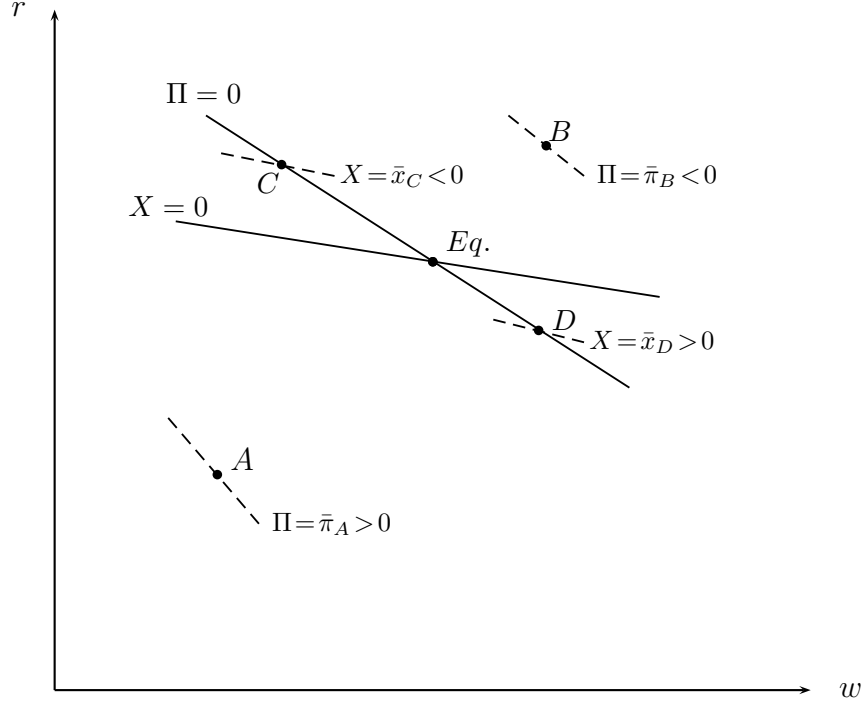


Figure C.2: Iso-profit and iso-excess demand curves

Notes: Note that we do not know the curvature of the two curves. Below the two solid zero-lines profit and excess demand are positive and above they are negative.

C.6 Intuition for convergence to stationary equity level

From the characteristics of the optimal contract (Figure 3.4), we notice that at low levels of promised values V^E expected repayments, $\pi_l m_l(V^E) + (1 - \pi_l) m_h(V^E)$, from entrepreneurs to banks exceed the level of bank loans $b(V^E)$ and that the opposite holds at high levels of promised values (see Figure C.3). Intuitively, this means that banks receive a positive net cash flow from entrepreneurs with low promised values. This positive net flow accrues to banks' equity. This is supplied as capital on the capital market and generates returns, which lead to a further accumulation of equity. In contrast, banks expect a negative net cash flow from firms with high promised values, which detracts banks' equity.

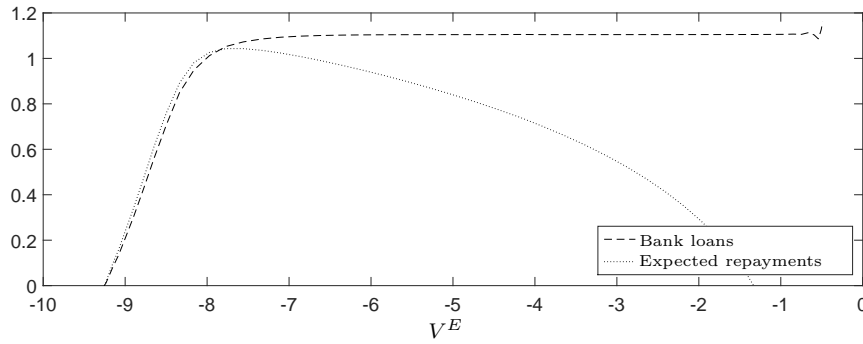


Figure C.3: Bank loans, $b(V^E)$, and expected repayments, $\pi_l m_l(V^E) + (1 - \pi_l)m_h(V^E)$

With this in mind, we can now intuitively describe the process of development of banks' equity level from the very beginning of time with no population to the stationary equity level E :⁴ Suppose the banks are endowed with E_0 at the beginning of time when there is no population in the economy, yet. As population starts, there is a new-born cohort of entrepreneurs (and workers) with promised values $V^W(0; r, w) = V_0^E$. Entrepreneurs sign contracts with banks, which entitle them to banks loans and which ask for repayments. At the beginning of their lives, when entrepreneurs are at low levels of promised values they must give positive net cash flows to banks. Hence, banks start accumulating equity. With age, the average promised value of entrepreneurs increases (see firm dynamics in Figure 3.7 and 3.8) and reaches eventually levels where banks loans are larger than expected repayments. This reduces banks' equity. In addition, as the economy evolves, there are more overlapping cohorts – with younger cohort making positive and older cohorts making negative net cash flows to banks. In aggregation there is an accumulation of total bank's equity. Finally, in the stationary equilibrium the accumulation of banks' equity come to a halt so that the equity level stays constant. This means, in equilibrium negative aggregate net payments from entrepreneurs are exactly covered by the interest generated on banks' equity.

⁴Assume for simplicity that during the process of development interest rate and wage are fixed at some level (e.g., the equilibrium level (r, w)).

C.7 Additional figures

C.7.1 Life paths

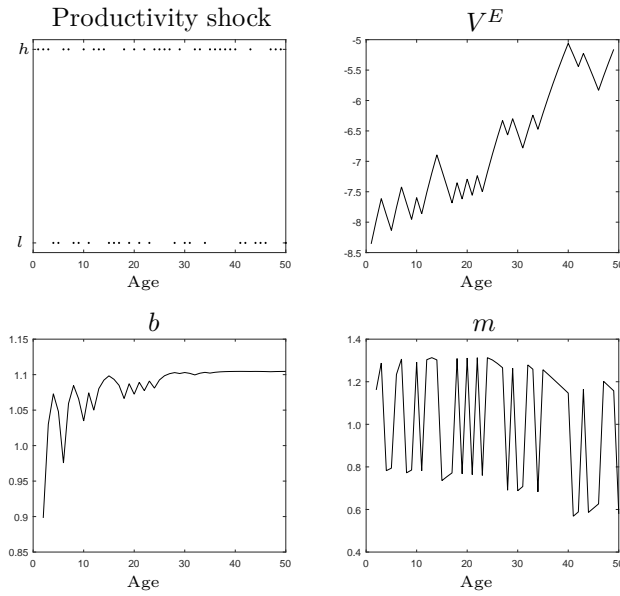


Figure C.4: Life path I

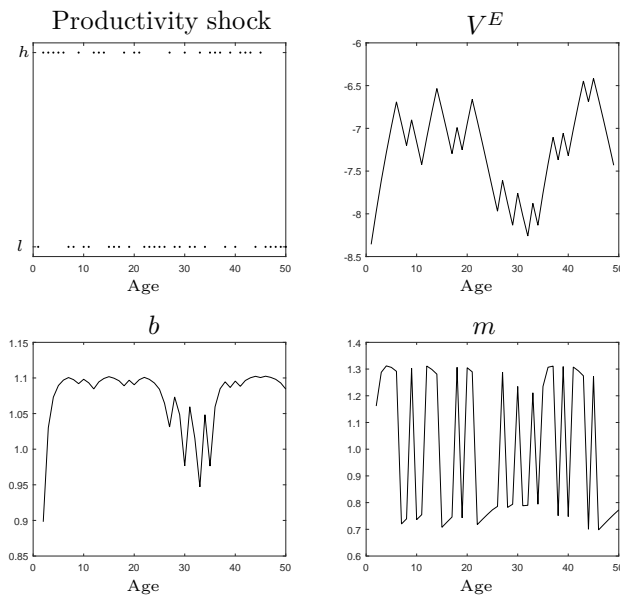


Figure C.5: Life path II

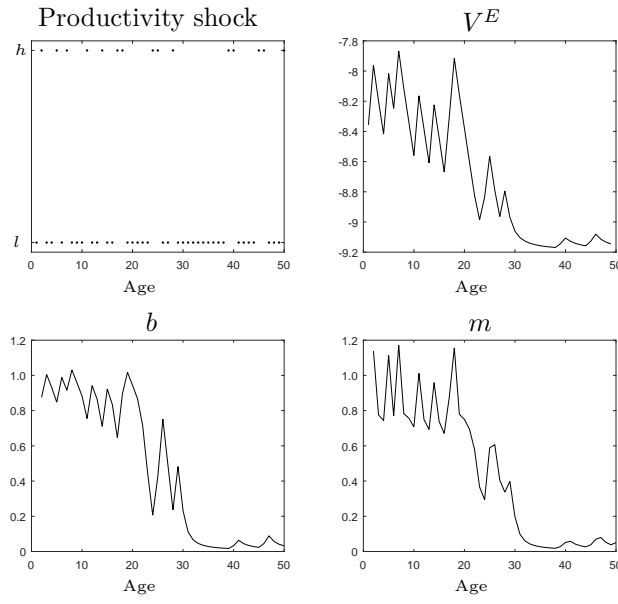


Figure C.6: Life path III

C.7.2 Development of entrepreneurs' distributions

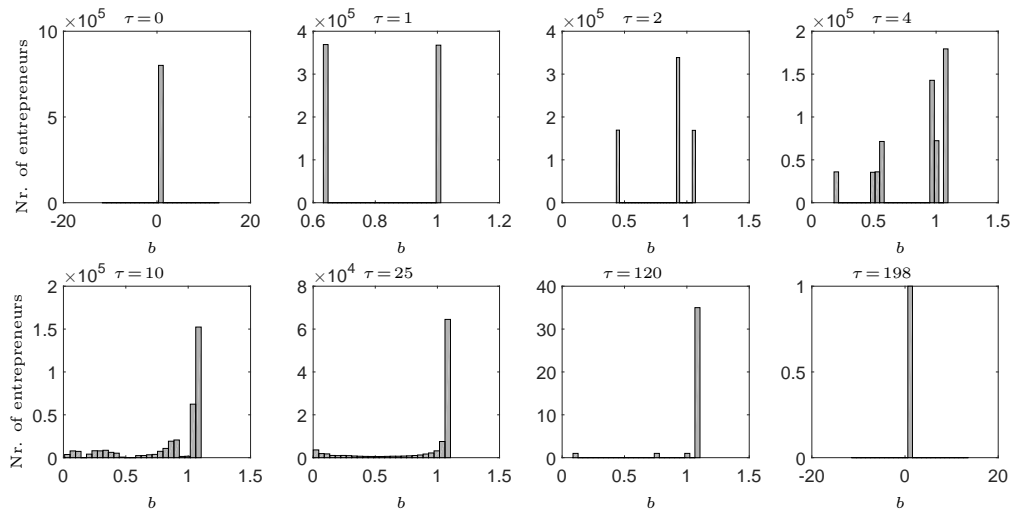


Figure C.7: Development of entrepreneurs' bank loans distributions

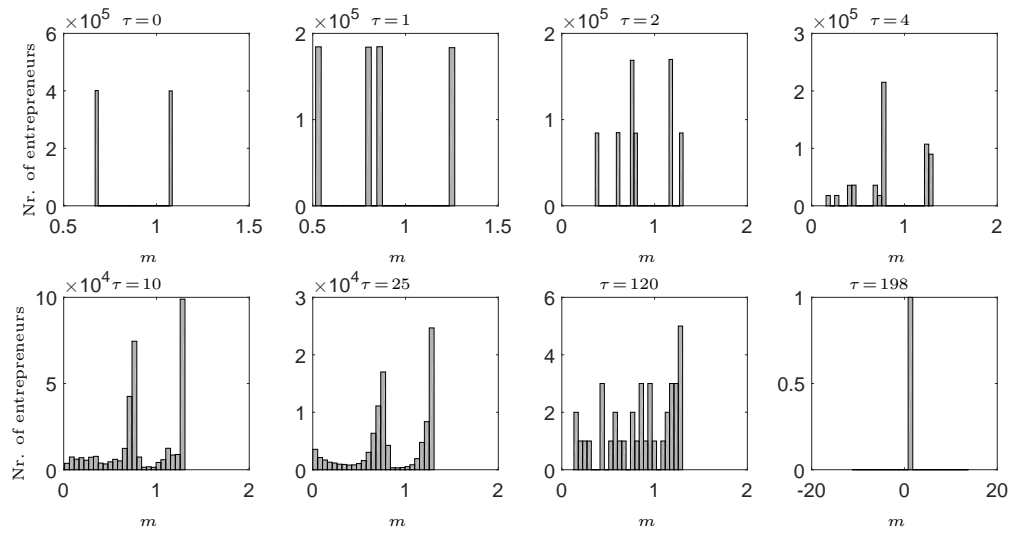


Figure C.8: Development of entrepreneurs' repayments distributions

Part IV

Bibliography

Bibliography

- Acemoglu, D. (2002). Directed technical change. *Review of Economic Studies* 69(4), 781–809.
- Acemoglu, D. and F. Zilibotti (1997). Was Prometheus unbound by chance? Risk, diversification, and growth. *Journal of Political Economy* 105(4), 709–751.
- Achury, C., S. Hubar, and C. Koulovatianos (2012). Saving rates and portfolio choice with subsistence consumption. *Review of Economic Dynamics* 15(1), 108–126.
- Admati, A. and M. Hellwig (2013). *The Bankers’ New Clothes: What’s Wrong with Banking and What to Do about It*. Princeton: Princeton University Press.
- Albuquerque, R. and H. A. Hopenhayn (2004). Optimal lending contracts and firm dynamics. *Review of Economic Studies* 71(2), 285–315.
- Allen, F. and D. Gale (1988). Optimal security design. *Review of Financial Studies* 1(3), 229–263.
- Allen, F. and D. Gale (1991). Arbitrage, short sales, and financial innovation. *Econometrica* 59(4), 1041–1068.
- Allen, F. and D. Gale (1994). *Financial innovation and risk sharing*. Cambridge, Massachusetts: MIT Press.
- Allen, F. and A. M. Santomero (1998). The theory of financial intermediation. *Journal of Banking & Finance* 21(11-12), 1461–1485.
- Angelini, P. and A. Generale (2008). On the evolution of firm size distributions. *The American Economic Review* 98(1), 426–438.
- Atkeson, A. and R. E. J. Lucas (1992). On efficient distribution with private information.

Review of Economic Studies 59(3), 427–435.

Atkeson, A. and R. E. J. Lucas (1995). Efficiency and equality in a simple model of efficient unemployment insurance. *Journal of Economic Theory* 66(1), 64–88.

Atkinson, A. B. (1997). Bringing income distribution in from the cold. *The Economic Journal* 107(441), 297–321.

Beck, T., B. Büyükkarabacak, F. K. Rioja, and N. T. Valev (2012). Who Gets the Credit? And Does It Matter? Household vs. Firm Lending Across Countries. *The B.E. Journal of Macroeconomics* 12(1), 1–44. Article 2.

Berger, A. N. and G. F. Udell (2002). Small business credit availability and relationship lending: The importance of bank organisational structure. *The Economic Journal* 112(477), F32–F53.

Bisin, A. (1998). General equilibrium with endogenously incomplete financial markets. *Journal of Economic Theory* 82(1), 19–45.

Boppart, T. (2014). Structural change and the Kaldor facts in a growth model with relative price effects and non-Gorman preferences. *Econometrica* 107(441), 297–321.

Boppart, T. (2015). To which extent is the rise in the skill premium explained by an income effect? *mimeo*.

Brunnermeier, M. K. and Y. Sannikov (2014). A macroeconomic model with a financial sector. *The American Economic Review* 104(2), 379–421.

Cabral, L. M. and J. Mata (2003). On the evolution of the firm size distribution: Facts and theory. *The American Economic Review* 93(4), 1075–1090.

Capelle-Blancard, G. and C. Labonne (2011). More Bankers, More Growth? Evidence from OECD Countries. *CEPII Document de Travail* 2011-22.

Carroll, C. D. (2002). Portfolios of the Rich. In L. Guiso, M. Haliassos, and T. Jappelli (Eds.), *Household Portfolios*. Cambridge, Massachusetts: MIT Press.

Carvalho, V. and X. Gabaix (2013). The great diversification and its undoing. *The*

- American Economic Review* 103(5), 1697–1727.
- Cecchetti, S. G. and E. Kharroubi (2015). Why does financial sector growth crowd out real economic growth? *BIS Working Papers* 490.
- Célérier, C. and B. Vallée (2014). The motives for financial complexity: An empirical investigation. *HEC Paris Research Paper No. FIN-2013-1013*.
- Célérier, C. and B. Vallée (2015). Returns to talent and the finance wage premium. *mimeo*.
- Clementi, G. L. and H. A. Hopenhayn (2006). A theory of financing constraints and firm dynamics. *Quarterly Journal of Economics* 121(1), 229–265.
- Cone, K. (1983). *Regulation of Depository Institutions*. Ph. D. thesis, Stanford.
- DeMarzo, P. M. and M. J. Fishman (2007). Agency and optimal investment dynamics. *The Review of Financial Studies* 20(1), 151–188.
- Diamond, P. A. (1967). The role of stock market in a general equilibrium model with technological uncertainty. *The American Economic Review* 57(4), 759–776.
- Duffie, D. and M. O. Jackson (1989). Optimal innovation of futures contracts. *The Review of Financial Studies* 2(3), 275–296.
- Duffie, D. and R. Rahi (1995). Financial market innovation and security design: An introduction. *Journal of Economic Theory* 65(1), 1–42.
- Dyrda, S. (2014). Fluctuations in uncertainty, efficient borrowing constraints and firm dynamics. *mimeo*.
- Epstein, G. (2005). Introduction: Financialization and the World Economy. In G. Epstein (Ed.), *Financialization and the World Economy*, Chapter 1, pp. 3–16. Cheltenham: Edward Elgar.
- European Central Bank ECB (accessed 08.07.2015). Statistical data warehouse: Number of listed securities. <https://sdw.ecb.europa.eu/browse.do?node=4212921>.
- Evans, D. S. (1987). The relationship between firm growth, size, and age: Estimates for

- 100 manufacturing industries. *The Journal of Industrial Economics* 35(4), 567–581.
- Eyster, E. and G. Weizsäcker (2011). Correlation neglect in financial decision-making. *Discussion papers // DIW Berlin German Institute for Economic Research* 1104.
- Falkinger, J. (2014). In search of economic reality under the veil of financial markets. *Working paper series / Department of Economics* 154.
- Falkinger, J., S. Studer, and Y. Zhao (2015). Explaining structural change towards and within the financial sector. *Working paper series / Department of Economics* 206.
- Federal Reserve Bank of St.Louis (accessed 12.08.2015). Effective federal funds rate. <https://research.stlouisfed.org/fred2/series/DFF>.
- Federation of European Securities Exchanges FESE (accessed 03.07.2015). Historical data. <http://www.fese.eu/statistics-market-research/historical-data>.
- Föllmi, R. and J. Zweimüller (2008). Structural change, Engel's consumption cycles and Kaldor's facts of economic growth. *Journal of Monetary Economics* 55(7), 1317–1328.
- Frame, W. S. and L. J. White (2004). Empirical studies of financial innovation: Lots of talk, little action? *Journal of Economic Literature* 42(1), 116–144.
- Freixas, X. and J.-C. Rochet (2008). *Microeconomics of Banking* (2nd ed.). Cambridge, Massachusetts: The MIT Press.
- Gennaioli, N., A. Shleifer, and R. W. Vishny (2012). Neglected risk, financial innovation, and financial fragility. *Journal of Financial Economics* 104(3), 452–468.
- Gennaioli, N., A. Shleifer, and R. W. Vishny (2013). A model of shadow banking. *The Journal of Finance* 68(4), 1331–1363.
- Gennaioli, N., A. Shleifer, and R. W. Vishny (2014). Finance and the preservation of wealth. *Quarterly Journal of Economics* 129(3), 1221–1254.
- Gorton, G. and A. Winton (2003). Financial intermediation. In G. M. Constantinides, M. Harris, and R. Stulz (Eds.), *Handbook of the Economics of Finance*, Volume 1(A) Corporate Finance, pp. 431–551. Amsterdam: Elsevier.

- Green, E. J. (1987). Lending and the smoothing of uninsurable income. In E. C. Prescott and N. Wallace (Eds.), *Contractual Arrangements for Intertemporal Trade*. University of Minnesota Press, Minneapolis.
- Greenwood, J. and B. Jovanovic (1990). Financial development, growth, and the distribution of income. *Journal of Political Economy* 98(5), 1076–1107.
- Greenwood, J. and D. Scharfstein (2013). The Growth of Finance. *Journal of Economic Perspectives* 27(2), 3–28.
- Gross, T. and S. Verani (2013). Financing constraints, firm dynamics, and international trade. *Finance and Economics Discussion Series 2013-02*. Washington: Board of Governors of the Federal Reserve System.
- Gründler, K. and J. Weitzel (2012). The financial sector and economic growth in a panel of countries. *Wirtschaftswissenschaftliche Beiträge des Lehrstuhls für Volkswirtschaftslehre, insbes. Wirtschaftsordnung und Sozialpolitik, Universität Würzburg Prof. Dr. Norbert Berthold* 123.
- Haliassos, M. (Ed.) (2013). *Financial Innovation: Too much or too little?* Cambridge, Massachusetts: MIT Press.
- Hall, B. H. (1987). The relationship between firm size and firm growth in the U.S. manufacturing sector. *The Journal of Industrial Economics* 35(4), 583–606.
- Jacklin, C. J. (1986). Demand deposits, trading restrictions, and risk sharing. In E. Prescott and N. Wallace (Eds.), *Contractual Arrangements for Intertemporal Trade*, pp. 26–47. Minneapolis, MN: University of Minnesota Press.
- Judd, K. L. (1998). *Numerical Methods in Economics*. Cambridge, Massachusetts: MIT Press.
- Kallir, I. and D. Sonsino (2009). The neglect of correlation in allocation decisions. *Southern Economic Journal* 75(4), 1045–1066.
- Kero, A. (2013). Banks' risk taking, financial innovation and macroeconomic risk. *The*

- Quarterly Review of Economics and Finance* 53(2), 112–124.
- King, M., S. Ruggles, J. T. Alexander, S. Flood, K. Genadek, M. B. Schroeder, B. Trampe, and R. Vick (2010). *Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. [Machine-readable database]*. Minneapolis: University of Minnesota.
- Kneer, C. (2013, September). The Absorption of Talent into Finance: Evidence from U.S. Banking Deregulation. *DNB Working Paper 391*.
- Law, S. H. and N. Singh (2014). Does too much finance harm economic growth. *Journal of Banking and Finance* 41, 36–44.
- Levine, R. (2005). Finance and growth: Theory and evidence. *Handbook of Economic Growth* 1A, 855–934.
- Levy, H. and H. M. Markowitz (1979). Approximating expected utility by a function of mean and variance. *The American Economic Review* 69, 308–317.
- Ljungqvist, L. and T. J. Sargent (2000). *Recursive macroeconomic theory*. Cambridge, Massachusetts: MIT Press.
- Machin, S. and J. Van Reenen (1998). Technology and changes in skill structure: Evidence from seven OECD countries. *The Quarterly Journal of Economics* 113(4), 1215–1244.
- Markowitz, H. M. (1952). Portfolio selection. *The Journal of Finance* 7(1), 77–91.
- Merton, R. C. (1969). Lifetime portfolio selection under uncertainty. *Review of Economics and Statistics* 51(3), 247–257.
- Miller, M. H. (1986). Financial innovation: The last twenty years and the next. *Journal of Financial and Quantitative Analysis* 21(4), 459–471.
- OECD Data (accessed 29.06.2015). Employment: Self-employment rate. <https://data.oecd.org/emp/self-employment-rate.htm>.
- OECD.Stat (accessed 07.06.2013). Finance. Bank Profitability Statistics. Income Statement and Balance Sheet. <http://stats.oecd.org/>.
- Pesendorfer, W. (1995). Financial innovation in a general equilibrium model. *Journal of*

Economic Theory 65(1), 79–116.

- Philippon, T. (2012). Notes on equilibrium financial intermediation. *mimeo NYU*.
- Philippon, T. (2015). Has the US Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation. *The American Economic Review* 105(4), 1408–1438.
- Philippon, T. and A. Reshef (2007). Skill biased financial development: Education, wages and occupations in the U.S. financial sector. *NBER Working Paper* 13437.
- Philippon, T. and A. Reshef (2012). Wages and human capital in the U.S. financial industry: 1909-2006. *Quarterly Journal of Economics* 127(4), 1551–1609.
- Philippon, T. and A. Reshef (2013). An international look at the growth of modern finance. *Journal of Economic Perspectives* 27(2), 73–96.
- Piketty, T. (2014). *Capital in the twenty-first century*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.
- Piketty, T. and E. Saez (2003). Income inequality in the United States, 1913-1998. *Quarterly Journal of Economics* 118(1), 1–39.
- Quadrini, V. (2004). Investment and liquidation in renegotiation-proof contracts with moral hazard. *Journal of Monetary Economics* 51(4), 713–751.
- Radner, R. (1985). Repeated principal-agent games with discounting. *Econometrica* 53(5), 1173–1198.
- Rogerson, W. P. (1985). Repeated moral hazard. *Econometrica* 53(1), 69–76.
- Ross, S. A. (1989). Institutional markets, financial marketing, and financial innovation. *The Journal of Finance* 44(3), 541–556.
- Rousseau, P. L. and P. Wachtel (2011). What is happening to the impact of financial deepening on economic growth? *Economic Inquiry* 49(1), 276–288.
- Samuelson, P. A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics* 51(3), 239–246.

- Shanghai Stock Exchange SSE (accessed 10.07.2015). Number of Listings. <http://english.sse.com.cn/information/statistics/historical/listings/>.
- Simsek, A. (2013a). Financial innovation and portfolio risk. *American Economic Review: Paper & Proceedings* 103(3), 198–401.
- Simsek, A. (2013b). Speculation and risk sharing with new financial assets. *The Quarterly Journal of Economics* 128(3), 1365–1396.
- Smith, A. A. J. and C. Wang (2006). Dynamic credit relationships in general equilibrium. *Journal of Monetary Economics* 53(4), 847–877.
- Spear, S. E. and S. Srivastava (1987). On repeated moral hazard with discounting. *The Review of Economic Studies*, 54(4), 599–617.
- Suellow, T. (2015). The skill-intensity of financial service consumption: An input-output analysis. *mimeo*.
- Swiss Exchange SIX (accessed 03.07.2015). Annual statistics. http://www.six-swiss-exchange.com/statistics/annual_statistics/.
- The Economist (2015). As safe as houses. *The Economist* 2015, January 31, 60.
- Thomas, J. and T. Worrall (1990). Income fluctuation and asymmetric information: An example of a repeated principal-agent problem. *Journal of Economic Theory* 51(2), 367–390.
- Tobin, J. (1985). Liquidity preferences as behaviour towards risk. *Review of Economic Statistics* 25(2), 65–86.
- Townsend, R. M. (1982). Optimal multiperiod contracts and the gain from enduring relationships under private information. *Journal of Political Economy* 90(6), 1166–1186.
- Tufano, P. (2003). Financial innovation. In G. M. Constantinides, M. Harris, and R. Stulz (Eds.), *Handbook of the Economics of Finance*, Volume 1(A) Corporate Finance, pp. 307–335. Amsterdam: Elsevier.

- U.S. Bureau of Economic Analysis (BEA) (accessed 08.10.2015). Gross-domestic-product-(gdp)-by-industry data. http://www.bea.gov/industry/gdpbyind_data.htm.
- U.S. Bureau of the Census (accessed 12.08.2015). Average Poverty Thresholds. <https://www.census.gov/hhes/www/poverty/data/historical/index.html>.
- Verani, S. (2015). Aggregate consequences of dynamic credit relationships. *Finance and Economics Discussion Series 2015-063*. Washington: Board of Governors of the Federal Reserve System, <http://dx.doi.org/10.17016/FEDS.2015.063>.
- World Bank (accessed 04.11.2015). World Development Indicators: Life expectancy at birth, total (years). <http://data.worldbank.org/indicator/SP.DYN.LE00.IN>.
- World Bank (accessed 05.11.2015). World Development Indicators: Gross savings (percentage of GDP). <http://data.worldbank.org/indicator/NY.GNS.ICTR.ZS>.

Part V

Curriculum Vitae

Curriculum Vitae

Personal details

Name: Sabrina Studer
Date of Birth: August, 16 1986
Place of Birth: St.Gallen, Switzerland
Nationality: Swiss

Education

09/2011 – 02/2016 PhD studies at the Zurich Graduate School of Economics
University of Zurich, Switzerland
Supervisor: Prof. Dr. Dr. Josef Falkinger
09/2009 – 07/2011 Master of Arts HSG in Quantitative Economics and Finance
University of St.Gallen, Switzerland
10/2005 – 07/2008 Bachelor of Arts HSG in Economics
University of St.Gallen, Switzerland

Professional experience

09/2011 – 01/2016 Research and teaching assistant at the Department of Economics,
University of Zurich
11/2009 – 09/2011 Teaching assistant at the Department of Economics,
University of St.Gallen
02/2009 – 04/2009 Internship at the KOF Swiss Economic Institute, ETH Zurich
08/2008 – 01/2009 Internship at the Swiss National Bank, Zurich
07/2005 – 07/2008 Internship and part-time job at Lupa Verkaufs AG, St.Gallen